

Theoretical Studies of the Dissociation of Charged Protein Complexes in the Gas Phase

by

Surajith Nalanth Wanasundara

B.Sc.(Hons), The University of Sri Jayewardenepura, Sri Lanka, 1998

M.Sc., The Memorial University of Newfoundland, Canada, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2009

© Surajith Nalanth Wanasundara 2009

Abstract

Understanding the dissociation mechanism of multimeric protein complex ions is important for interpreting gas-phase experiments. The aim of this thesis work was to study the dissociation of charged protein complexes. To pursue this, a number of model and molecular dynamics calculations were conducted using the cytochrome *c'* dimer.

The energetics of differing charge states, partitionings, and configurations were examined in both the low and high charge regimes. It is shown that one must always consider distributions of charge configurations, once protein relaxation effects are taken into account, and that no single configuration dominates. These results also indicate that in the high charge limit, the dissociation is governed by electrostatic repulsion from the net charges. This causes two main trends: i) charges will move so as to approximately maintain constant surface charge density, and ii) the lowest barrier to dissociation is the one that produces fragment ions with equal charges.

Free energies are also calculated for the protonated dimer ion as a function of the center of mass distance between the monomers. In addition, the change of intermolecular properties such as intermolecular hydrogen bonds and the smallest separation of intermolecular residues were analyzed. It is found that monomer unfolding competes with complex dissociation, and that the relative importance of these two factors depends upon the charge partitioning in the complex. Symmetric charge partitionings preferentially suppress the dissociation barrier relative to unfolding, and complexes tend to dissociate promptly with little structural change occurring in the monomers. Alternatively, asymmetric charge partitionings preferentially lower the barrier for monomer unfolding relative to the dissociation barrier. In this case, the monomer with the higher charge unfolds before the complex dissociates. For large multimeric proteins, the unfolding and subsequent charging of a single monomer is a favorable process, cooperatively lowering both the unfolding and dissociation barriers at the same time. For the homodimer considered here, this pathway has a large free energy barrier.

Overall, the work presented herein demonstrates that molecular dynamics simulation can be useful for understanding the dissociation mechanism of protein complexes.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgements	ix
Dedication	x
1 Introduction	1
1.1 Using ESI-MS to Study Non-covalent Protein Complexes	2
1.2 Dissociation of Protein Complexes in the Gas Phase	3
1.3 Charge Partitioning During the Dissociation of Protein Complexes	4
1.4 Theoretical Models of Asymmetric Charge Partitioning	6
1.5 Molecular Dynamics Simulations of Biomolecules	8
1.6 Objectives	9
2 Theory	10
2.1 Molecular Dynamics Simulation	10
2.1.1 Empirical Force Fields	10
2.1.2 Newtonian Equations and Numerical Integration	12
2.1.3 Berendsen Thermostat for Constant Temperature MD	14
2.1.4 Bond Constraint	15
2.2 Free Energy Calculations	17
2.3 Protein Shape Descriptors	20
2.3.1 Mean Number of Overcrossings	20
2.3.2 Radius of Gyration	21

3	Potential Energy Calculations	22
3.1	Bare Charge Model	22
3.1.1	Method	22
3.1.2	Results	25
3.2	Empirical Potential Energy Calculations	26
3.2.1	MD Simulations	26
3.2.2	Results	28
3.3	Discussion	34
3.3.1	Coulomb Repulsion Model	34
3.3.2	Application of the Coulomb Repulsion Model	36
4	Dissociation Barrier Estimation	42
4.1	Method	42
4.2	Results	44
4.3	Discussion	51
5	Dissociation Coordinates	56
5.1	Method	56
5.2	Results	56
5.3	Discussion	62
6	Concluding Remarks	66
	References	68
 Appendices		
A	Charge Sites	78
B	OPLS Atom Types for the Heme Prosthetic Group of Cytochrome c'	79
C	List of Publications	80
D	Corrections for the Pull Code in Gromacs 3.2	81

List of Tables

3.1	Calculated gas-phase protonation energies of the basic amino acids using the MP2/6-311+G(d,p) method.	25
3.2	Average mean number of overcrossings taken at 1 ps and 20 ps for MD runs of a sample of 50 charge configurations selected with the lowest energy from the distributions in Figures 3.3 and 3.4. *Chains 1 and 2 denote the monomer with the higher and lower charges, respectively.	32
5.1	For differing charge partitionings, this table compares the average distance that the last intermolecular hydrogen bond between two monomers breaks, R_{hb} , with the COM position of the peak of the corresponding calculated free energy barrier from Figure 4.2.	58

List of Figures

2.1	Basic MD algorithm.	13
3.1	Crystal structure of the cytochrome c' dimer.	23
3.2	Relative potential energies (filled circles - read against the left vertical scale) and intermolecular potential energies (open squares - read against the right vertical scale) of the lowest energy charge configuration as a function of charge partitioning calculated using the bare charge method for the (a) D10 and (b) D18 total charge states. Here M_x/M_y denotes x charges on one monomer and y charges on the other.	27
3.3	Distributions of relative potential energies for short run trajectories started with three groups of initial conditions, denoted "low", "medium", and "high", taken from the bare charge calculation. These results are for the M7/M3 charge partitioning of the D10 charge state.	29
3.4	The same as Figure 3.3 except for the M10/M8 charge partitioning of the D18 charge state.	29
3.5	Averages of the relative potential energies at 1 ps (circles) and 20 ps (squares) of a MD run of an ensemble of 50 charge configurations selected with the lowest energy from the distributions in Figures 3.3 and 3.4, as a function of charge partitioning between two monomers in the (a) D10 and (b) D18 total charge states. Here M_x/M_y denotes x charges on one monomer and y charges on the other. The error bars indicate the widths of the distributions of the average energy of the ensemble.	31
3.6	Snapshots of cytochrome c' structures initially (a) and after a 20 ps MD simulation with the (b) M6/M4 and (c) M13/M5 charge partitioning. In each complex, the higher and lower charged monomers are drawn to the left and right respectively.	34
4.1	Constraint force as a function of center of mass distance between two monomers for the (a) M5/M5 (b) M6/M4 (c) M7/M3 (d) M8/M2 and (e) M9/M1 charge partitions. These are each averages over 50 different trajectories. The errors bar indicate the standard deviation of the average values.	45

4.2	Relative free energy changes as a function of COM distance for different charge partitionings. Here M_x/M_y denotes x charges on one monomer and y charges on the other.	46
4.3	Average mean number of overcrossings as a function of COM distance for the (a) M5/M5 (b) M6/M4 (c) M7/M3 (d) M8/M2 and (e) M9/M1 charge partitionings. Vertical bars indicate the positions of the barriers in Figure 4.2. In each panel, the upper and lower curves correspond to values for the monomer with the smaller and larger charges, respectively.	48
4.4	Average radius of gyration as a function of COM distance for the (a) M5/M5 (b) M6/M4 (c) M7/M3 (d) M8/M2 and (e) M9/M1 charge partitionings. Vertical bars indicate the positions of the barriers in Fig. 4.2. In each panel, the lower and upper curves correspond to values for the monomer with the smaller and larger charges, respectively.	49
4.5	Snapshots from the constraint MD run at a COM distance of 6 nm for the (a) M5/M5 and (b) M8/M2 charge partitionings (the monomer with the higher charge is shown on the left).	50
4.6	Schematic two dimensional free energy landscape as a function of complex dissociation and monomer unfolding coordinates for (left panel) a symmetric charge partitioning and (right panel) an asymmetric charge partitioning. The dashed lines represent possible paths for traversing from the bound dimer state, at the origin, to larger center of mass distances.	52
5.1	The average number of intermolecular hydrogen bonds as a function of COM distance for five different trajectories started from the same charge configuration in the (a) M5/M5 and (b) M7/M3 charge partitionings.	57
5.2	Average R_{rd} distances as a function of COM distance for the (a) Neutral, (b) M5/M5, (c) M7/M3 and (d) M8/M2 charge partitionings (solid line). Bars indicate the standard deviation of the average. The dashed blue vertical lines are at the corresponding R_{hb} values from Table 5.1. The red dashed lines represent the idealized change of R_{rd} when two monomers separate without structural changes (slope =1).	61
5.3	Changes in the potential energy components (angle bending (black), dihedral angle (red), intramolecular Lennard-Jones (green), intramolecular Coulomb (blue), intermolecular Lennard-Jones (violet), intermolecular Coulomb (brown)) and total potential energy (orange) as a function of COM distance for the (a) neutral, (b) M5/M5, (c) M7/M3 and (d) M8/M2 charge partitionings. Vertical dashed lines indicate the corresponding values of R_{hb} from Table 5.1.	63

A.1	Amino acid sequence of the cytochrome <i>c'</i> monomer. Charge sites are represented by bold letters.	78
B.1	OPLS atom types for the heme prosthetic group. Hydrogen atoms are also included in parameter file and not shown here for clarity	79

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Mark Thachuk, for encouragement, invaluable guidance and support throughout every aspect of my Ph.D. program and research. His advice and support in the editing of this thesis is also greatly appreciated.

I also want to thank my thesis committee members, especially Dr. Don Douglas for his careful reading of my thesis. Special thanks are extended to Dr. Roman Baranowski and Mr. Brent Gawryluik for giving support in technical issues of computing on the WestGrid computing facilities.

I owe loving thanks to my wife Anusha. Without her love, encouragement and understanding, it would have been impossible for me to finish this work. My special gratitude is due to my parents, brothers and sisters for their loving support and encouragement.

I truly thank faculty members and my fellow graduate students in Theoretical Chemistry who have made my stay here at UBC enjoyable and meaningful.

I acknowledge financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada. All computations were performed using WestGrid computing resources, which are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions (www.westgrid.ca).

*This thesis is dedicated
to
my parents*

Chapter 1

Introduction

Electrospray ionization (ESI) mass spectrometry (MS) has been widely used to study large biomolecules as well as non-covalent biomolecular complexes [1–5]. Electrospray ionization allows the relatively gentle phase transfer of biomolecules complexes from solution to the gas phase. Furthermore, the nature of the soft transition in the ESI process not only avoids fragmentation of intact protein complexes, but also often preserves weak non-covalent interactions [3, 5, 6]. Thus, electrospray ionization with mass spectrometry is a powerful tool to study non-covalent biomolecular complexes.

Particularly, tandem mass spectrometry (MS/MS) can be a useful tool to determine structural information of large non-covalent protein complexes by dissociating selected gaseous ions [7, 8]. Unfortunately, the interpretation of tandem mass spectra of these complexes is hindered by a relatively poor understanding of the dissociation mechanism [9]. Many groups [6, 10–18] have reported an asymmetric dissociation pattern for large multi-meric proteins complexes as a function of charge to mass ratio. A small subunit, typically a protein monomer, is ejected from the complex during dissociation, with the monomer carrying away a disproportionate amount of charge for its relative mass. It is necessary to understand the origin of the asymmetric dissociation behavior of proteins complexes if gas-phase dissociation experiments are to be successfully used to obtain structural information such as the relative stability of complexes in solution and the binding topology of complexes. Even though a number of models have been proposed to explain the charge partitioning among fragment ions during dissociation of protein complexes [6, 10, 12, 18–20], the phenomenon is not fully understood. All these reported models have used either simple electrostatic models or static protein/complex structures, which did not include the dynamics of the proteins. Molecular dynamics simulation, which allows structural relaxations to occur as a function of charge state, will be a useful tool in examining the charge partitioning process.

This chapter provides a general introduction to electrospray ionization mass spectrometry (ESI-MS). An overview of the asymmetric charge distributions obtained from the dissociation of charged protein complexes in ESI-MS experiments, and models developed to explain this phenomenon are described. As most of the work in this thesis is performed using molecular dynamics (MD) simulation techniques, a brief introduction of MD is also given. Overview and objectives of this study are given at the end.

1.1 Using ESI-MS to Study Non-covalent Protein Complexes

Mass spectrometry is an analytical technique that is based upon the motion of charged particles in an electric or a magnetic field. A mass spectrometer consists of an ion source that generates gas-phase ions, a mass analyzer that separates the ionized analytes according to the mass-to-charge (m/z) ratio, and a detector that measure abundances of ions at each m/z value.

ESI [21–23] is the most commonly used method to create ions of intact protein and protein complexes for MS studies [7, 24]. ESI gently transfers non-covalent complexes in solution into the gas phase allowing the study of those complexes by mass spectrometry. This technique also produces highly charged assemblies, thereby lowering the mass-to-charge ratio enough to allow detection of larger intact proteins and protein complexes. In typical ESI, highly charged droplets are produced by forcing the analyte solution (sample) through a thin metal capillary tip. A sufficiently high potential (2 - 3 KV) is used at the end of the metal capillary to disperse the solution into a very fine spray of charged droplets. The solvent evaporates, reducing the droplet diameter and increasing the charge concentration at the droplet’s surface. Eventually, at a given radius (Rayleigh limit), Coulombic repulsion overcomes the droplet’s surface tension and the droplet breaks up into smaller droplets. This Coulombic explosion forms a series of smaller charged droplets [2]. The process is repeated until individually charged naked analyte ions are formed by either solvent evaporation from single droplets (charged residue model) [21] or ejection from the droplet surface (ion evaporation model) [25].

Matrix assisted laser desorption ionization (MALDI) [26, 27] is also a soft technique that enables the transfer of intact proteins into the gas phase without fragmentation [7, 24]. However, the sample preparation requirement of MALDI is not an ideal environment for keeping proteins under physiological conditions, and most likely denatures the protein due to dried and often highly acidic conditions in the sample [7]. Moreover, MALDI is limited in the low-mass range because of matrix-associated chemical noise [28].

The mass analyzer is used to separate ions within a selected range of mass-to-charge (m/z) ratios. The analyzer is an important part of the instrument because of the role it plays in the instrument’s accuracy and mass range. Ions are typically separated by magnetic fields, electric fields, or by measuring the time it takes an ion to travel a fixed distance. Currently, there are a number of mass analyzers available, the better known of which include magnetic sectors, quadrupole analyzer [29], Time-of-Flight (ToF) analyzer [30], quadrupole ion trap [31] and Fourier transform ion cyclotron resonance (FT-ICR) analyzer [32].

1.2 Dissociation of Protein Complexes in the Gas Phase

Tandem mass spectrometry (MS/MS) has emerged as an important technique to obtain structural information about protein complexes. This technique involves dissociation of selected precursor ions and mass analysis of products. In general, ions of a single m/z (precursor ions) formed in the ion source are mass-selected by a first mass analyzer (MS^1) and focused into a region where they are activated in some way that causes them to fall apart to produce fragment (product) ions. Product ions then proceed into a second mass analyzer (MS^2) for analysis. The ion activation/dissociation step plays a major role in these experiments. Several ion activation techniques have been developed to study large protein complexes and some of the commonly used methods are described below.

Presently, collision-induced (or collisionally activated) dissociation (CID) is the most commonly used activation method. In this method, precursor ions are accelerated and allowed to collide with a neutral target gas. Each collision converts part of the translational energy into internal energy of the ion. Repeated collisions build up internal energy in the ion, until eventually the fragmentation threshold is reached and product ions are formed [17, 33]. The CID process can be categorized into two groups based on the translational energy of precursor ions: low-energy collisions and high-energy collisions [34]. Low-energy collisions occur in the range of 1-100 eV collision energy and are common in the quadrupole ion trap, triple quadrupole (QqQ) and FT-ICR instruments. High-energy collisions are in the keV range and are common in sector and ToF instruments, which is achieved by placing a collision cell between two mass analyzers in both instruments [34].

Surface-induced dissociation (SID) involves colliding accelerated mass-selected precursor ions with a surface instead of a neutral gas molecule. Upon collision with the surface, some of the ion translational energy is converted to internal energy, with resulting activation of the precursor ion and subsequent fragmentation [35]. SID is a fast, single step activation method. Internal energy is deposited into the precursor ion within picoseconds [9]. Several different instruments, such as sector, FT-ICR, ToF, and quadrupole ion trap, have been employed for SID experiments [36].

Blackbody infrared radiative dissociation (BIRD) is a low energy thermal activation method. In this technique, ions are thermally excited until dissociation occurs. This is achieved by a constant exchange of energy between ions and their surroundings due to absorption and emission of infrared photons, which are always present in an environment in the form of the blackbody radiation field [12, 37, 38]. The BIRD method is usually performed in the FT-ICR instrument. There are two essential requirements for BIRD experiments: (i) extremely low ambient pressure (below 10^{-6} Torr) that allows IR photon absorption and emission to compete significantly with collisional energy exchange, and (ii) long observation times (on the order of seconds) which gives enough time to observe any

significant degree of dissociation from this relatively slow process [38].

The sustained off-resonance irradiation collisionally activated dissociation (SORI-CAD) technique is also a slow heating mechanism and available with the FT-ICR instruments. In the SORI-CAD technique, precursor ions are excited by a pulse with slightly off-resonance frequency, causing the ions' kinetic energy to oscillate with time that allows the ions to undergo multiple low-energy collisions with an inert target gas. As the ion collides with the target gas, its internal energy slowly increases. When sufficient energy transfers from translation to internal energy, ion fragmentation occurs [39].

The electron capture dissociation (ECD) method is currently unique to FT-ICR MS [40]. During ECD, low-energy electrons are injected into the ICR cell and captured by multiply charged precursor ions, leading to charge state reduction and subsequent fragmentation of ions [41].

1.3 Charge Partitioning During the Dissociation of Protein Complexes

A number of research groups have investigated the dissociation of a variety of non-covalent protein complexes. Results from several of these studies indicate a highly asymmetric dissociation with respect to the mass to charge ratio with a number of dissociation/activation methods [6, 11–14, 16–18]. For example, SORI-CAD experiments conducted by Smith and co-workers [6] showed that 14+ streptavidin tetramer predominantly produces +7 monomer/+7 trimer and +6 monomer /+8 trimer ion pairs. Results were explained by speculating that the dissociation of the tetramer may occur by a Coulombically driven process in which a monomer species becomes unraveled and is ejected from the aggregate with a disproportionately large share of charge [10].

Klassen and coworkers [12, 18] have conducted a series of gas-phase dissociation studies, mostly on Shiga like toxin pentamer and streptavidin tetramer, by employing the BIRD technique. They observed that dissociation of the complex proceeds almost exclusively by loss of a single monomer unit with a disproportionately large fraction of the total charge. In these BIRD experiments, the Arrhenius activation energy (E_a) and pre-exponential factor (A) have been determined by measuring the temperature dependence of dissociation rate constants. They found that large Arrhenius A factors are ranging from a modest 10^{16} to extremely large 10^{39} , indicating that loss of the subunit is entropically favorable which corresponds to a large transition state entropy. The charge enrichment of the leaving subunit is expected to unfold the subunit and break the inter-subunit interactions, thus, increasing the number of low-frequency internal modes and the transition state entropy. The authors proposed that charge enrichment of the leaving subunit is energetically unfavorable but

destabilizes the complex sufficiently by increasing the repulsion between the subunit and the rest of the complex as well as by Coulombic repulsion-induced unfolding of the leaving subunit [12].

Benesch *et al.* [17] investigated the correlation between the relative surface areas of the products and division of the charge during dissociation. In this study, ions of the 12-mer of *TaHSP16.9* and 24-mer of *MjHSP16.5* were activated through collisions with argon atoms. The typical dissociation pattern was ejection of a monomer subunit carrying away a large percentage of the charge of the original complex. Furthermore, it shows that the partitioning of charge between the monomers and resulting oligomers closely follows the ratios of their estimated surface areas. Here, the ejected monomer is assumed to be in an unfolded state. For example, the monomer resulting from the dissociation of *MjHSP16.5* had 30% of the surface area with respect to the total surface area of the fragment ions and carried 29% of the precursor oligomer’s charge [17]. Chowdhury *et al.* [42] pointed out that multiply charged ions are produced primarily as a result of proton attachment to the available basic sites in the protein, and that the availability of ionizable basic sites depends upon the conformation of the protein in solution. In general, a protein in an unfolded conformation may possess more available basic sites than those in tightly folded conformations.

Recently, a few studies have shown exceptions to the exclusive loss of highly charged monomer subunits from gas-phase dissociation of an intact protein oligomer [35, 43, 44]. SID experiments conducted by Wysocki and coworkers [35] have shown that multimeric complexes dissociate into equally charged monomeric parts. For example, a tetramer ejects a monomer carrying away approximately 25% of the total charge. SID induces protein complex dissociation quickly compared with the timescales for other motions. Furthermore, the rapid deposition of higher internal energy allows better access to dissociation pathways that may lead to charge symmetric dissociation [35].

Heck and coworkers [43] have studied dissociation of 2-keto-3- deoxyarabinonate dehydratase and arabinose dehydrogenase which are stable as tetramers of similar size. 2-keto-3- deoxyarabinonate dehydratase complexes dissociate almost exclusively into two dimeric fragment ions, with near symmetric charge partitioning. Based on the crystal structure, the authors suggested that the nature of the intersubunit contacts were responsible for this uncommon dissociation pattern. In other words, 2-keto-3- deoxyarabinonate dehydratase complexes are composed of dimers that have more binding interactions internally than between them, making them more prone to dissociate into dimeric fragments. However, tetrameric arabinose dehydrogenase complexes, stabilized by extensive intersubunit interactions among monomers, dissociates by producing highly charged monomeric subunits and trimeric species, the typical dissociation pattern observed by other groups [43]. Furthermore, ECD of gp31 oligomer studied by the same group exhibits a main dissociation pathway resulting in a hexamer and monomer, the charge separation over the two products

is highly proportional to molecular weight [44]. ECD causes rapid disassembly of a complex and the leaving subunit does not unfold during dissociation from the complex [44].

Some reports indicate that asymmetric charge separation occurs even during homodimer dissociation under some experimental conditions [9, 19, 33, 45]. Gas-phase CID studies were carried out by Heck and co-workers for dimers of *E. coli* Glyoxalase I, human galectin, horse heart cytochrome c, and hen egg lysozyme [33]. They found that the charge distribution over two monomers of both Glyoxalase I and cytochrome c were highly asymmetric while that of both human galectin and hen egg lysozyme were a mix of symmetric and asymmetric.

Using Fourier-transform mass spectrometry (FTMS), Jurchen and Williams [19] conducted a study in order to understand the factors which may influence charge separation during protein dimer dissociation. In these experiments, isolated charge states of cytochrome c dimer were dissociated by SORI-CAD. According to Jurchen and Williams [19], asymmetric charge distribution depends upon charge state, dissociation energy, and conformational flexibility. These studies showed that higher charge states lead to symmetric charge products while lower charge states lead to an asymmetric charge dissociation pathway. Furthermore, their results with different excitation energies show that symmetric charge partitioning occurs when ions are activated using low energies while an asymmetric charge partitioning occurs at higher energies. They also observed that reducing the conformational flexibility of the proteins decreases the extent of asymmetric dissociation of the complex. These results suggested that the origin of asymmetric charge partitioning in these homodimers is the result of one of the protein monomers unfolding in the dissociation transition state allowing it to carry away the vast majority of charge [19].

1.4 Theoretical Models of Asymmetric Charge Partitioning

Several theoretical models have been developed to explain charge partitioning among fragment ions after protein complex dissociation. Smith and co-workers [6] employed the charged droplet model (CDM) of Ryce and Wyman [46] to explain highly asymmetric charge partitioning during their SORI-CAD experiment with tetrameric ions. In CDM, the protein complex is treated as a sphere. Assuming that the electric charge resides uniformly on the surface of the drops, the energy of the system at a “pseudo saddle point” is approximated as

$$E = 4\pi\gamma (r_1^2 + r_2^2) + k_e \frac{q_1^2}{2r_1} + k_e \frac{q_2^2}{2r_2} + k_e \frac{q_1 q_2}{\alpha(r_1 + r_2)} , \quad (1.1)$$

in which the first term on the right of the equation is the surface energy and the other terms represent the electric energy. Here, q_1 and q_2 are net surface charges on two separated drops of radii r_1 and r_2 , respectively, γ is surface tension, α is a scale factor between 1 and ∞ that defines the point at which the two drops are formed, and k_e is the Coulomb force constant.

For an example, $\alpha = 1$ if dissociation occurs when the two spheres are just touching. The energy is minimized when

$$0 = \frac{\partial E}{\partial q_1} = k_e \left\{ \frac{q_1}{r_1} + \frac{q_2}{r_2} \frac{\partial q_2}{\partial q_1} + \frac{1}{\alpha(r_1 + r_2)} \left(q_1 \frac{\partial q_2}{\partial q_1} + q_2 \right) \right\} . \quad (1.2)$$

Since $q_1 + q_2 = Q(\text{total charge})$, $\frac{\partial q_2}{\partial q_1} = -1$. Thus, Eq. 1.2 is simplified to

$$\frac{q_1}{r_1} - \frac{q_2}{r_2} + \frac{q_2 - q_1}{\alpha(r_1 + r_2)} = 0 , \quad (1.3)$$

and the charge ratio of two charged droplets is given by

$$\frac{q_1}{q_2} = \frac{r_1}{r_2} \left[\frac{\alpha r_1 + (\alpha - 1)r_2}{\alpha r_2 + (\alpha - 1)r_1} \right] = \frac{r_1}{r_2} \left[\frac{\alpha \frac{r_1}{r_2} + (\alpha - 1)}{\alpha + (\alpha - 1)\frac{r_1}{r_2}} \right] . \quad (1.4)$$

Smith and co-workers [6] assume that mass is proportional to volume. Therefore, the ratio between radii of monomer and trimer is given by $r_1/r_2 = (m_1/m_2)^{1/3} = 0.693$. Substituting this result into Eq. 1.4 gives the ratio of $(m_1/z_1)/(m_2/z_2)$ to vary between 0.48 for $\alpha = \infty$ and 0.69 for $\alpha = 1$. Here, (m_1/z_1) and (m_2/z_2) are charge-to-mass ratios for a monomer and trimer, respectively. The ratio stays less than 1 which implies that smaller fragments should get more than their share of charges. When Heck and co-workers [33] applied the CDM to their homodimeric dissociations, it showed that equal mass fragments would produce equally charged fragments. Heck and co-workers concluded that the CDM could not quantitatively account for their results.

Csiszar and Thachuk [20] studied charge distributions using the discretely charged ellipsoid model (DCEM). In this model, the shape of protein monomers was approximated by ellipsoids, both of the prolate and oblate varieties. A number of sizes, shapes, orientations, and types of ellipsoids were considered with several combinations of randomly generated charge sites on both ellipsoid surfaces. For each case, electrostatic energies as well as two charge transfer parameters have been calculated as a function of the fractional surface area. The results from this study showed that charge asymmetry depends upon the relative surface area of the monomers, with charges distributing themselves to keep constant surface charge density.

Furthermore, Klassen and coworkers [18] performed a number of model calculations, also based upon electrostatic calculations of charges distributed both on idealized spheres and on protein structures. In their studies, the lowest energy charge state of a leaving subunit (monomer) was determined for the protein complex structures with folded, partially unfolded, and fully unfolded states of leaving subunits. Calculations were also conducted with continuous and discrete charged droplet models. They have compared calculated val-

ues with values obtained from BIRD-CAD experiments [18]. They draw conclusions that are generally consistent with previous studies, and found that their measured charge distributions by BIRD experiments can be predicted by either a simple discretely charged sphere model or by a more detailed model that includes actual protein structures incorporating monomers with varying degrees of unfolding. In other words, the charge distributions are qualitatively consistent with the surface area ratios of the product ions.

1.5 Molecular Dynamics Simulations of Biomolecules

Molecular dynamics (MD) simulations have been used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. Performing MD simulations helps to get information which is difficult or impossible to obtain from experiments. In other words, MD simulations play an increasingly prominent role in understanding and interpreting experiments at the microscopic level as well as in studying regions which are not accessible experimentally, or which would imply very expensive experiments, such as under extremely high pressure. The availability of a number of general purpose computer simulation programs such as Amber [47], Charmm [48, 49], Gromacs [50–52], Gromos [53] and Namd [54] have made the MD simulation technique an attractive tool for research. Moreover, increasing computational power and continuing advances in methodology have extended MD studies to larger systems, greater conformational changes, and longer time scales. Currently, it is possible to perform MD simulations with the aim of studying microscopic properties of systems involving tens of thousands of atoms with timescales of nanoseconds and beyond [55].

The MD simulation technique was introduced and applied by Alder and Wainwright in the late 1950’s to study phase transitions in hard-sphere liquids [56], and has developed over the last six decades. Followed by a number of realistic simulations of simple systems [57], the first MD simulation of a macromolecule was published in 1977 with the simulation of the bovine pancreatic trypsin inhibitor (BPTI) [58]. Since then, MD has been used to study the dynamics of large macromolecules, including biological systems such as proteins [57], nucleic acids (deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)) [59], carbohydrates [60] and phospholipids [61].

Atomistic level MD simulations consist of the numerical solving of the classical equation of motion of a system using a suitable interaction potential. Interactions of all atoms in a given system are described by a relatively simple empirical potential function or force field. The quality of the results of a MD simulation depend on the accuracy of the potential. The forces on all atoms can be calculated and integrated in time, using time steps on the order of 10^{-15} s. The primary result is a trajectory, which contains the motion of all atoms in time, over millions of time steps. Analysis of these motions gives insight into the system

that is being studied.

1.6 Objectives

The goal of this thesis work is to investigate using a model system how charge separation during gas-phase dissociation of non-covalent multimeric protein complexes can occur and the factors that influence it. Experimental studies evidence that asymmetric charge distributions during the dissociation of protein complexes is not unique to any specific protein complex [6, 9, 11–14, 16–19, 33, 45]. Thus, it is expected that this phenomenon will not depend sensitively upon specific protein interactions. Any reasonable model system should be adequate for understanding the qualitative behavior of the dissociation process. Performing MD simulations with large protein complexes is too time consuming. On the other hand, small cytochrome c dimers have been studied by dissociation experiments [9, 19, 33] and performing MD simulations with cytochrome c dimer is computationally viable. Cytochrome *c'* dimer, which is one type of cytochrome c obtained from the purple phototrophic bacterium *Chromatium vinosum* [62], was therefore used as a model for the current study. Furthermore, Cytochrome *c'* exists as a dimer and each monomer has enough basic sites to protonate for the current study.

Three different aspects as described below were studied and discussed in this thesis.

- i. The energetics of differing charge states, charge partitionings, and charge configurations were examined in both the low and high charge regimes. Simplified charge calculations, as well as finer, atomistic level semi-empirical calculations using molecular dynamics were performed for the cytochrome *c'* dimer in the gas phase (Chapter 3).
- ii. Free energy profiles were then calculated to gain insight into the dissociation mechanism. More specifically, constrained molecular dynamics (MD) calculations were used to estimate the free energy changes as a function of the distance between the centers of mass of two monomers in a dimeric complex (Chapter 4).
- iii. Finally, dissociation properties such as the number of intermolecular hydrogen bonds and intermolecular residual distances were studied as a function of the distance between centers of masses of two monomers. In addition to the analysis of these parameters, components of the potential energy function were also studied (Chapter 5).

Chapter 2

Theory

2.1 Molecular Dynamics Simulation

2.1.1 Empirical Force Fields

In classical MD simulation, interactions between particles are described by a force field that is a combination of a relatively simple potential function and parameters which are used in that function. The success of MD simulation often depends on the quality of the force field. Furthermore, ranges of applicability and the accuracies of these force fields have been limited by the description of the potential function as well as the way parameters are derived. Parameters are most often derived based on experimental data or high level quantum mechanical data of small model compounds [63–65]. Over the past decades, a number of biomolecular force fields have been developed and improved by several research groups [66]. Amber [63, 64], Charmm [48, 67], Gromos [53], and OPLS [65, 68] are some of the commonly used force fields.

The classical potential energy function is defined as a function of the coordinates of each of the atoms. The function is comprised of covalent and non-covalent terms.

Covalent Potential Energy

The covalent interactions are typically described by a sum of bond stretching (1-2 interactions), angle bending (1-3 interactions) and torsion rotation (1-4 interactions) terms. Bond stretching and angle bending are usually treated harmonically. The potential energies of bond stretching and angle bending are therefore given by respectively,

$$V_{bond} = \sum_{i=1}^{N_{bond}} \frac{1}{2} k_i^{bond} (r_i - r_{eq,i})^2 , \quad (2.1)$$

$$V_{angle} = \sum_{i=1}^{N_{angle}} \frac{1}{2} k_i^{angle} (\theta_i - \theta_{eq,i})^2 . \quad (2.2)$$

Here r , θ , k^{bond} and k^{angle} are the bond length, bond angle and force constants for bond stretching and angle bending, respectively. The subscript i is used to represent the i^{th} bond or i^{th} angle and the subscript eq represents the equilibrium values. Equilibrium bond

lengths, angles and values for the force constants are determined from crystal structures and by fitting to vibrational frequency data for model compounds [63, 67]. *Ab initio* data from small reference molecules are also introduced to supplement the experimental geometric data [67].

The dihedral or torsional rotations are modeled as a truncated Fourier series. The OPLS-AA/L force field uses the first three terms in the expansion [65] and the potential energy of torsional rotations is given by

$$V_{dihedral} = \sum_{i=1}^{N_{dihedral}} \left[\frac{V_1^i}{2} [1 + \cos(\phi^i - \gamma_1^i)] + \frac{V_2^i}{2} [1 - \cos(2\phi^i - \gamma_2^i)] + \frac{V_3^i}{2} [1 + \cos(3\phi^i - \gamma_3^i)] \right], \quad (2.3)$$

in which ϕ^i is the dihedral angle, V_1^i , V_2^i and V_3^i are the coefficients in the Fourier series, and γ_1^i , γ_2^i and γ_3^i are phase angles. Here the superscript i indicates the i^{th} dihedral angle. The phase angle determines where the dihedral angle passes through its minimum value. Some force fields such as Charmm [48, 67] include only the first term in the series. Parameters for the torsional term are obtained from experimental data. Furthermore, these values are adjusted to match torsional barriers extracted from experiments or from *ab initio* calculations [63, 65]. In addition to bond stretching, angle bending and torsion rotation terms, additional covalent terms are often included in the force field. For example, improper dihedral angles, based on a harmonic term, are used to treat out-of-plane bending motions [67].

Non-covalent Potential Energy

The non-covalent potential energy corresponds to interactions between atoms in different molecules as well as between atoms in the same molecule separated by three or more covalent bonds. These interactions consist of electrostatic and Lennard-Jones two-body interaction terms. The electrostatic interactions are described by the Coulomb interaction with static, partial atomic charges. The Lennard-Jones potential is a combination of attractive van der Waals forces due to induced dipole-dipole interactions and empirical repulsive forces due to non-bonded overlap between the electron clouds and the nuclear repulsion at very short internuclear separations. Thus, the non-covalent potential energy term is given by

$$V_{NB} = \sum_{\substack{\text{non-bonded} \\ \text{atom pairs } ij}} \left[\frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{(12)}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{(6)}}{r_{ij}^6} \right) \right] f_{ij}, \quad (2.4)$$

in which r_{ij} is the atomic distance between particles i and j , ϵ_0 is the dielectric constant in vacuum and ϵ_r is the relative dielectric constant. Here, q_i and q_j are partial charges on particles i and j , respectively. The Lennard-Jones parameters ϵ_{ij} , $\sigma_{ij}^{(12)}$ and $\sigma_{ij}^{(6)}$ depend on pairs of atom types and are obtained from combining rules: $\epsilon_{ij}^{(12)} = \left(\epsilon_{ii}^{(12)} \times \epsilon_{jj}^{(12)}\right)^{1/2}$, $\epsilon_{ij}^{(6)} = \left(\epsilon_{ii}^{(6)} \times \epsilon_{jj}^{(6)}\right)^{1/2}$ and $\sigma_{ij} = (\sigma_{ii}\sigma_{jj})^{1/2}$. The scale factor f_{ij} is 1.0 for atom pairs in the same molecule separated by more than three covalent bonds as well as for two atoms in two different molecules. Those non-covalent interactions separated by exactly three bonds (1-4 interactions) are reduced by the application of the scale factor which permits the use of the same parameters for inter and intramolecular interactions. For example, f_{ij} in the OPLS-AA/L force field is 0.5 for those 1-4 atoms pairs [65].

There are two common charge determination methods for biomolecular force fields. One is *ab initio* electrostatic potential (ESP) approaches where the partial charges are obtained by fitting the electrostatic potentials estimated on a grid surrounding the molecule by using HF/6-31G level *ab initio* calculations [69]. In the second method, partial atomic charges are assigned to a model compound to reproduce properties of organic liquids [70]. Lennard-Jones parameters are determined from scattering, crystal packing or liquid crystal data as well as quantum mechanics calculations.

2.1.2 Newtonian Equations and Numerical Integration

In MD simulations, atoms move according to Newton's law. The classical equations of motion are given by

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\{\mathbf{r}_i\}) , \quad i = 1, 2, 3, \dots, N , \quad (2.5)$$

in which m , \mathbf{r} , and \mathbf{F} are mass, position vector, and force, respectively. Here the subscript i is used to represent the i^{th} particle in a N particle system. The force \mathbf{F}_i can be derived from the interparticle potential energy V using

$$\mathbf{F}_i(\{\mathbf{r}_i\}) = -\frac{\partial V(\{\mathbf{r}_i\})}{\partial \mathbf{r}_i} . \quad (2.6)$$

Once the potential energy as a function of atom positions is known, given the coordinates of a starting structure and a set of velocities, the force acting on each atom is calculated and a new set of coordinates is generated numerically by solving the classical equation of motion for each particle. Before a MD run starts, the initial coordinates and velocities of all particles are required. If initial velocities are not available, they are usually generated according to a Maxwell-Boltzmann distribution at a desired temperature. The basic MD algorithm is shown in Figure 2.1.

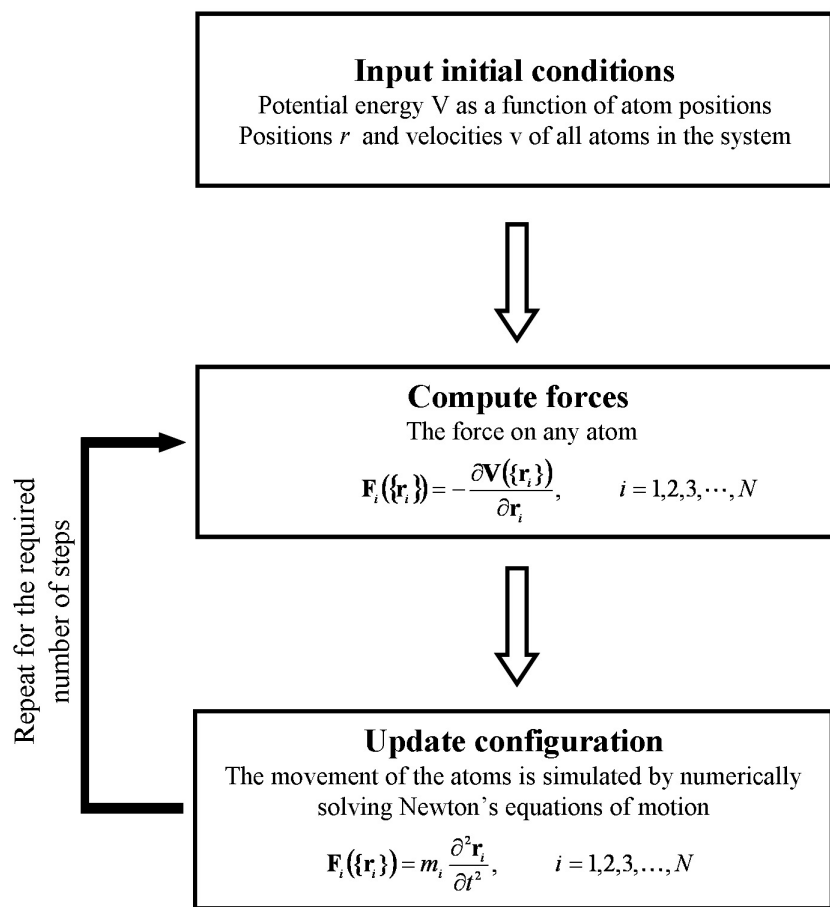


Figure 2.1: Basic MD algorithm.

Many numerical algorithms have been developed for integrating the equations of motion. The Leapfrog algorithm [71], which is used in the current study, is one of the commonly used algorithms in MD simulations and is described below.

Velocities of particle i , \mathbf{v}_i , at time point $t + \frac{\Delta t}{2}$ and $t - \frac{\Delta t}{2}$ can be approximated by a Taylor series expansion from time point t , that is

$$\mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) = \mathbf{v}_i(t) + \frac{d\mathbf{v}_i(t)}{dt} \frac{\Delta t}{2} + \frac{1}{2} \frac{d^2\mathbf{v}_i(t)}{dt^2} \left(\frac{\Delta t}{2} \right)^2 + \dots \quad (2.7)$$

$$\mathbf{v}_i \left(t - \frac{\Delta t}{2} \right) = \mathbf{v}_i(t) - \frac{d\mathbf{v}_i(t)}{dt} \frac{\Delta t}{2} + \frac{1}{2} \frac{d^2\mathbf{v}_i(t)}{dt^2} \left(\frac{\Delta t}{2} \right)^2 - \dots \quad (2.8)$$

Subtracting Eq. (2.8) from Eq. (2.7) and rearranging gives

$$\mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) = \mathbf{v}_i \left(t - \frac{\Delta t}{2} \right) + \frac{d\mathbf{v}_i(t)}{dt} \Delta t + \dots \quad (2.9)$$

$$\mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) \approx \mathbf{v}_i \left(t - \frac{\Delta t}{2} \right) + \frac{\mathbf{F}_i}{m_i} \Delta t. \quad (2.10)$$

Using the same procedure for the Taylor expansion of \mathbf{r}_i from the time point $t + \frac{\Delta t}{2}$ we get

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \frac{d\mathbf{r}_i(t)}{dt} \Delta t + \dots \quad (2.11)$$

$$\mathbf{r}_i(t + \Delta t) \approx \mathbf{r}_i(t) + \mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) \Delta t. \quad (2.12)$$

Equations (2.12) and (2.10) are implemented in the Leapfrog algorithm. In this algorithm, velocities are first calculated at time $t + \frac{\Delta t}{2}$ and used to calculate the positions of the particles at time $t + \Delta t$ and so on. If velocities at $t + \Delta t$ are needed, they can be approximated by

$$\mathbf{v}_i(t + \Delta t) \approx \frac{1}{2} \left(\mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) + \mathbf{v}_i \left(t + \frac{3\Delta t}{2} \right) \right). \quad (2.13)$$

2.1.3 Berendsen Thermostat for Constant Temperature MD

Various methods have been developed to constrain the temperature in MD simulations. The Berendsen thermostat [72] and the Nosé–Hoover thermostat [73–75] are the most widely used methods. The Berendsen thermostat was used in this study because it has been found to be very efficient in equilibrating to the desired temperature and this permitted us to sample more trajectories by reducing the computational time in each MD run.

Velocity scaling of each atom at each time-step is the simplest way to fix the system temperature to a desired value T_0 . The velocity scaling factor κ can be determined as follows [76]. The time average of kinetic energy for an N particle system confined to volume

V at temperature T, $\langle K \rangle_{NVT}$, is given by

$$\langle K \rangle_{NVT} = \frac{3}{2} N k_B T = \sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i^2 , \quad (2.14)$$

in which k_B is the Boltzmann constant. Here, m_i and \mathbf{v}_i are the mass and velocity of particle i . If the temperature at time-step t is $T(t)$, the temperature change after the velocity scaling becomes

$$T_0 - T(t) = \sum_{i=1}^N \frac{1}{3} \frac{m_i (\kappa \mathbf{v}_i)^2}{N k_B} - \sum_{i=1}^N \frac{1}{3} \frac{m_i (\mathbf{v}_i)^2}{N k_B} = (\kappa^2 - 1) T(t) , \quad (2.15)$$

giving $\kappa = \sqrt{\frac{T_0}{T(t)}}$.

Berendsen *et al.* [72] modified the method by introducing an external heat bath which acts as a source of thermal energy, supplying or removing heat from the system as necessary. In this technique, velocities are scaled at each time step, such that the rate of change of temperature is proportional to the difference in temperature between the bath (T_0) and the system ($T(t)$),

$$\frac{dT(t)}{dt} = \frac{1}{\tau} (T_0 - T(t)) , \quad (2.16)$$

in which τ is the time constant which determines the rate of temperature scaling. Thus the change in temperature between successive time steps is given by

$$\Delta T = \frac{\Delta t}{\tau} (T_0 - T(t)) . \quad (2.17)$$

Equations (2.15) and (2.17) give the scaling factor for velocities as

$$\kappa = \left[1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T(t)} - 1 \right) \right]^{\frac{1}{2}} . \quad (2.18)$$

It should be noted that the Berendsen thermostat does not necessarily produce a canonical distribution [72] but present results should be satisfactory because only energy averages not energy fluctuations are calculated.

2.1.4 Bond Constraint

It is quite common practice to constrain intramolecular bonds in classical MD simulations, having fixed bond lengths [77]. Because very high frequency is associated with bond vi-

brations, application of bond-length constraints permits reduced computational effort by increasing the simulation time step [78]. Classical equations of motion for atoms in a specific molecule with distance constraints can be written as

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = - \frac{\partial}{\partial \mathbf{r}_i} \left[V(\{\mathbf{r}_i\}) + \sum_{k=1}^{N_c} l_k(t) \sigma_k(\{\mathbf{r}_i\}) \right], \quad k = 1, 2, \dots, N_c, \quad (2.19)$$

in which N_c is the number of constraints within the molecule and $l_k(t)$ is the Lagrange multiplier that enforces the k^{th} constraint,

$$\sigma_k(\{\mathbf{r}_i\}) = \mathbf{r}_{k_1 k_2}^2 - d_{k_1 k_2}^2 = 0, \quad k = 1, 2, \dots, N_c. \quad (2.20)$$

Here $\mathbf{r}_{k_1 k_2} = \mathbf{r}_{k_1} - \mathbf{r}_{k_2}$ is the link vector between the atoms k_1 and k_2 involved in constraint k , and $d_{k_1 k_2}$ is the corresponding constraint distance. The first term on the right-hand side of Eq. (2.19) represents the unconstrained force \mathbf{F}_i^{uc} acting on atom i , while the second term represents the constraint force \mathbf{F}_i^c and gives

$$\mathbf{F}_i^c(t) = - \sum_{k=1}^{N_c} l_k(t) \frac{\partial \sigma_k(\{\mathbf{r}_i\})}{\partial \mathbf{r}_i} = -2 \sum_{k=1}^{N_c} l_k(t) (\delta_{i, k_1} - \delta_{i, k_2}) \mathbf{r}_{k_1 k_2}(t). \quad (2.21)$$

The unconstrained position at time step $t + \Delta t$ (see Eqs. (2.10) and (2.12)) is given by the Leapfrog algorithm as

$$\mathbf{r}_i^{uc}(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i \left(t - \frac{\Delta t}{2} \right) + (\Delta t)^2 \frac{\mathbf{F}_i^{uc}(t)}{m_i}. \quad (2.22)$$

After applying constraint forces, the new position is given by

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i^{uc}(t + \Delta t) + (\Delta t)^2 \frac{\mathbf{F}_i^c(t)}{m_i}. \quad (2.23)$$

The positions of atoms k_1 and k_2 should satisfy the constraint of Eq. (2.20), and thus

$$\left\{ \mathbf{r}_{k_1}^{uc}(t + \Delta t) - \mathbf{r}_{k_2}^{uc}(t + \Delta t) + (\Delta t)^2 \frac{\mathbf{F}_{k_1}^c(t)}{m_{k_1}} - (\Delta t)^2 \frac{\mathbf{F}_{k_2}^c(t)}{m_{k_2}} \right\}^2 - d_{k_1 k_2}^2 = 0. \quad (2.24)$$

Substituting the expansion in Eq. (2.21) for the constraint forces into Eq. (2.24) leads to a quadratic equation, namely

$$\left\{ \mathbf{r}_{k_1}^{uc}(t + \Delta t) - \mathbf{r}_{k_2}^{uc}(t + \Delta t) - 2(\Delta t)^2 \frac{1}{m_{k_1}} \sum_{k'=1}^{N_c} l_{k'}(t) \left(\delta_{k_1, k'_1} - \delta_{k_1, k'_2} \right) \mathbf{r}_{k'_1 k'_2}(t) \right. \\ \left. + 2(\Delta t)^2 \frac{1}{m_{k_2}} \sum_{k'=1}^{N_c} l_{k'}(t) \left(\delta_{k_2, k'_1} - \delta_{k_2, k'_2} \right) \mathbf{r}_{k'_1 k'_2}(t) \right\}^2 - d_{k_1 k_2}^2 = 0 . \quad (2.25)$$

The set of N_c quadratic equations can be written by considering all constraints and are to be solved for the Lagrange multipliers. The Shake algorithm is the most commonly used method for solving this set of equations [79]. In the Shake method, two approximations are made in order to calculate the solutions: (1) the equations are linearized by neglecting the quadratic terms in the Lagrange multipliers, and (2) the equations are decoupled by considering that each constraint is isolated and not involved in any other constraint. Approximate solutions for the Lagrange multipliers are thereby given as,

$$l_k(t) = \frac{[\mathbf{r}_{k_1 k_2}^{uc}(t + \Delta t)]^2 - d_{k_1 k_2}^2}{4(\Delta t)^2 \left(m_{k_1}^{-1} + m_{k_2}^{-1} \right) \mathbf{r}_{k_1 k_2}(t) \cdot \mathbf{r}_{k_1 k_2}^{uc}(t + \Delta t)} , \quad k = 1, 2, \dots, N_c . \quad (2.26)$$

Both positions of atoms k_1 and k_2 are updated according to the following equations,

$$\mathbf{r}_{k_1}(t + \Delta t) = \mathbf{r}_{k_1}^{uc}(t + \Delta t) - \frac{2(\Delta t)^2}{m_{k_1}} l_k(t) \mathbf{r}_{k_1 k_2}(t) \quad \text{and} \\ \mathbf{r}_{k_2}(t + \Delta t) = \mathbf{r}_{k_2}^{uc}(t + \Delta t) + \frac{2(\Delta t)^2}{m_{k_2}} l_k(t) \mathbf{r}_{k_1 k_2}(t) . \quad (2.27)$$

An iteration procedure is required to solve for the Lagrange multipliers because Eq. (2.26) is not exact due to the two approximations made. The coordinates r_i obtained by Eq. (2.27) at each iteration are used as the coordinate r_i^{uc} for the next iteration. Iterations for all of the constraints of the system are cycled until all constraints satisfy the condition,

$$\frac{|r_{k_1 k_2}(t + \Delta t) - d_{k_1 k_2}|}{d_{k_1 k_2}} \leq \tau , \quad k = 1, 2, \dots, N_c , \quad (2.28)$$

in which τ is a given tolerance.

2.2 Free Energy Calculations

There are several approaches for determining free energy changes using molecular dynamic simulations [80]. Among them, thermodynamics integration and umbrella sampling with

the weighted histogram analysis method (WHAM) are two widely used methods [81]. In umbrella sampling, a biasing potential is added to the Hamiltonian to direct the simulations toward a certain goal. The biasing potential is usually a harmonic potential that keeps the reaction coordinate near a specified value. This is done in a number of windows along the reaction coordinate. WHAM is then used to determine the optimal free energy constants for the combined simulations. Calculated free energy constants are used to obtain the free energy profile along the reaction path.

The slow growth method, which was used in this study to calculate free energy difference, is derived from the thermodynamic integration technique [82, 83]. In this method, constraint MD simulations are performed by controlling the change of a predefined reaction coordinate. Even though this forces the reaction along a predefined reaction path, it is possible to estimate a free energy profile of the reaction with the choice of suitable reaction coordinates [84]. Such a calculation requires a knowledge of the Jacobian for the coordinate transformation between cartesian and generalized (reaction) coordinates [85]. In the case of dissociation of protein complexes, the distance between the centers of mass (COM) of the two dissociating fragments can be applied straightforwardly as a dissociation coordinate. In the present case, the distance between the centers of mass can be constrained using the Shake algorithm hence the more computationally intensive umbrella sampling method is not necessary.

The Helmholtz free energy of a system is defined as

$$F(\lambda) = -k_B T \ln Z(\lambda) , \quad (2.29)$$

in which the canonical (NVT) partition function of the system is given by

$$Z(\lambda) = \int_{\Gamma_{all}} \exp \left(-\frac{H(\lambda)}{k_B T} \right) d\Gamma . \quad (2.30)$$

The integration is carried out over all phase space Γ . The Hamiltonian $H(\lambda)$ is a function of some generalized coordinate λ in addition to the positions and momenta of particles. The free energy difference between initial state λ_1 and λ_2 can be calculated as

$$\Delta F = \int_{\lambda_1}^{\lambda_2} \frac{\partial F(\lambda)}{\partial \lambda} d\lambda . \quad (2.31)$$

From Eqs. (2.29) and (2.30), the derivative of $Z(\lambda)$ respect with λ can be obtained as

$$\begin{aligned}
\frac{\partial F(\lambda)}{\partial \lambda} &= -k_B T \frac{1}{Z(\lambda)} \frac{\partial Z(\lambda)}{\partial \lambda} \\
&= \frac{\int_{\Gamma_{all}} \left[\frac{\partial H(\lambda)}{\partial \lambda} \right] \exp \left(\frac{-H(\lambda)}{k_B T} \right) d\Gamma}{Z(\lambda)} \\
&= \left\langle \frac{\partial H_\lambda}{\partial \lambda} \right\rangle_\lambda,
\end{aligned} \tag{2.32}$$

in which $\langle \dots \rangle_\lambda$ denotes an equilibrium average over the intermediate state λ . Accordingly, the free energy difference between two states can be written as an integration of the derivative of the classical Hamiltonian, H , with respect to the reaction coordinate of interest λ as

$$\Delta F = \int_{\lambda_1}^{\lambda_2} \left\langle \frac{\partial H_\lambda}{\partial \lambda} \right\rangle_\lambda d\lambda. \tag{2.33}$$

Here $\partial H_\lambda / \partial \lambda$ represents the driving force for the quasistatic process along the coordinate λ . This equation provides a fundamental relationship to calculate free energy differences using equilibrium simulations. In other words, if the system is changed in an infinitely slow reversible process along the reaction path between two states, the work done on the system, W_{eq} , is equal to the free energy change of the system.

In practice, integration is determined numerically through intermediate equilibrium MD simulations at discrete points over the interval λ_1 and λ_2 . In such a case, each of the MD simulations should equilibrate in order to determine ensemble averages at each λ . However, if the equilibrium average is poorly calculated, such as can result when λ is changed too quickly, W will exceed the free energy, the excess representing dissipative work induced by the irreversible change in the system. For large systems, such as protein complexes, it is often the case that the sampling of phase space is insufficient to approach a reversible process, and hence dissipative work is present in calculations.

However, Jarzynski [86] has shown that free energy differences can be rigorously related to the work generated by nonequilibrium processes by

$$\exp(-\beta \Delta F) = \overline{\exp(-\beta W)}. \tag{2.34}$$

Unfortunately, in practice it can be difficult to converge the average on the right hand side because only those paths with the lowest work contribute significantly, and these can be of statistically small measure. One way of overcoming this difficulty is to assume a particular form for the distribution of work related to the irreversible processes, and then calculate the parameters that determine this form. Hummer [87] has shown that by representing the

distribution as a Gaussian, the free energy change can be estimated as

$$\Delta F \approx \overline{W} - \frac{\beta \sigma^2}{2}, \quad (2.35)$$

in which \overline{W} is the average nonequilibrium work taken over many measurements (that is, different trajectories in moving from one value of λ to another), $\sigma^2 = \overline{W^2} - \overline{W}^2$, and $\beta = 1/k_B T$. The second term on the right hand side gives an estimate of the dissipative work, and with this correction, the difference in free energy between two states can be approximated. Ultimately, in applying this equation one needs to obtain the average and deviation of the work done from a series of separate simulations, even if each is not in the reversible limit.

2.3 Protein Shape Descriptors

It is important to capture the secondary structural changes of proteins during the MD simulations. However, there is no unique classical or quantum mechanical property that characterizes the secondary structure of the protein. A variety of geometrical procedures have been developed to capture the change of secondary structure during MD simulations [88]. Two descriptors were used in this study to characterize the structural changes and are explained in detail below.

2.3.1 Mean Number of Overcrossings

The mean number of overcrossings is a simple geometrical descriptor to capture secondary structural changes in proteins. Consider counting the number of bonds that cross each other in a two dimensional projection of a protein structure. If one averages this number of bond-bond overcrossings over all possible projections, the mean number of overcrossings is obtained. It should be noted that the analysis is restricted to only main chain (backbone) bonds. The mean overcrossing number is given by [89],

$$\overline{N} = \lim_{m \rightarrow \infty} \sum_{N=0}^{\max N} N \frac{m_N}{m}, \quad (2.36)$$

in which m is the number of random bond-bond overcrossing projections with m_N projections containing N overcrossings. Thus, a decrease in this mean number indicates that less overcrossing occurs that is, the secondary structure of the protein is becoming less twisted.

2.3.2 Radius of Gyration

The radius of gyration is defined as the spatial distribution of atoms and does not depend upon connectivity and bonding within a molecule. The radius of gyration R_g defines the radius of the smallest sphere having a center at the center of mass of a molecule and completely enclosing all the n atoms and is given by,

$$R_g^2 = \left(\frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i} \right) , \quad (2.37)$$

in which m_i is the mass of the atom i and r_i is the position of atom i respect to the center of mass of the molecule.

Chapter 3

Potential Energy Calculations*

Potential energies were calculated in order to determine the favorable charge partitioning for the ground state dimer in both the low and high total charge regimes. These calculations were carried out for gas-phase ions of the cytochrome *c'* dimer using a simple electrostatic model (bare charged model) as well as a more detailed semi-empirical based MD simulation. Throughout this chapter, the term “charge partitioning” refers to the number of charges that are assigned to each monomer in a complex ion. The term “charge configuration” refers to the particular arrangement of charges on charge sites.

3.1 Bare Charge Model

3.1.1 Method

This calculation used a simple electrostatic model in which the electrostatic repulsion between the charges in the protein are the only interactions included. X-ray crystallographic data for the cytochrome *c'* dimer (PDB ID # 1bbh) from the purple phototrophic bacterium *Chromatium vinosum* (CVCP) were obtained from the Protein Databank (PDB) [62]. The cytochrome *c'* dimer structure is shown in Figure 3.1. This cytochrome *c'* dimer is composed of two monomers with identical subunits of 131 residues each. Each monomer consists of four antiparallel α -helices with a heme prosthetic group in the center. The specificity of the dimer interface is due to hydrophobic residues and not to any charged residues or intermolecular hydrogen bonds [62]. It should be noted that all water molecules in the crystal structure were removed for this study in order to mimic the complex in the gas phase where no water molecules remain after the ESI process. In this study, a simple approach with bare charges was used to choose the lowest energy charge distributions.

Protonation is assumed to occur only at basic sites that include arginine, histidine, lysine and the N-terminus. Covey et al. [90] found that the total number of basic sites was often similar to the maximum charge state of a protein. Smith et al. [1] also provided supporting details for this concept by compiling a list of proteins and peptides by including the maximum charge state obtained from ESI spectra and the number of basic sites. Using

*Results presented in this chapter appear in Wanasundara, S. N.; Thachuk, M., Theoretical investigations of the dissociation of charged protein complexes in the gas phase, J. Am. Soc. Mass Spectrom., 2007, 18, 2242-2253.

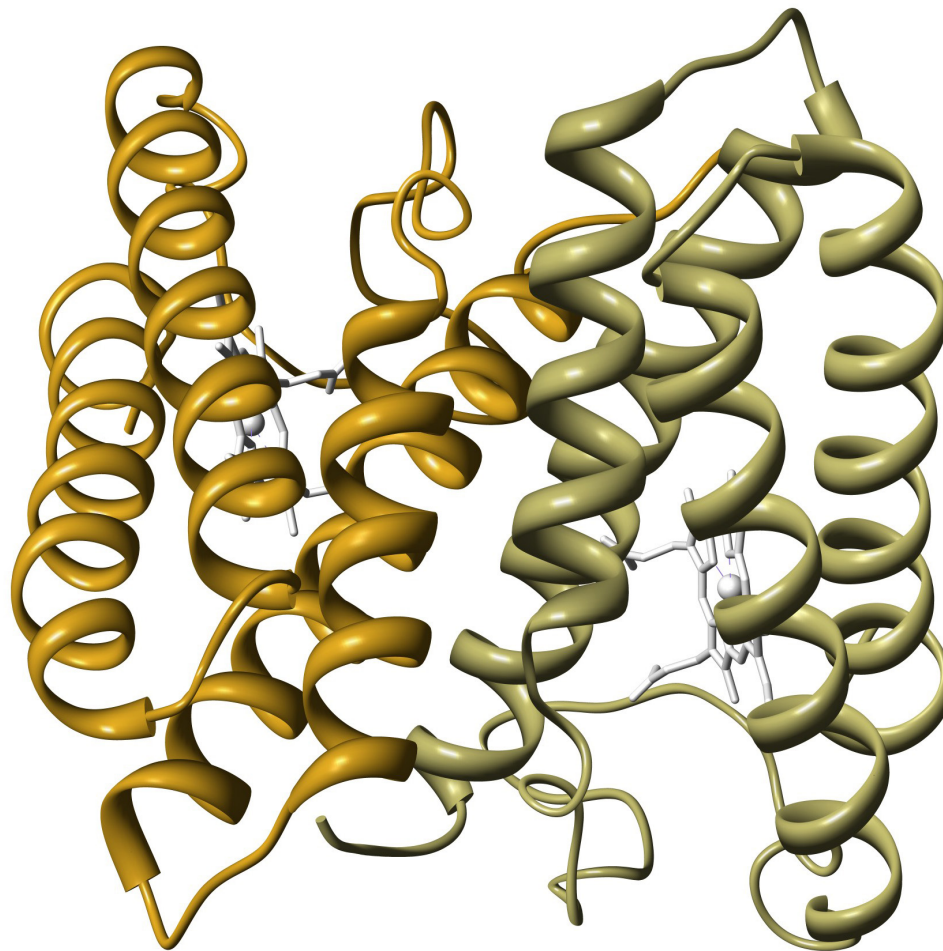


Figure 3.1: Crystal structure of the cytochrome c' dimer.

the same clarification, the cytochrome *c'* dimer has fifteen available basic sites on each of the two monomers (see Appendix A). There is an enormous number of ways of distributing protons among the available basic sites for each charge state. Therefore, performing molecular dynamics simulations with all possible charge permutations is not practical. To overcome this problem, a screening process must be used to choose a manageable number of charge configurations. Miteva et al. [91] used a procedure based on the Metropolis algorithm for Monte Carlo sampling of charge configurations. Schnier et al. [92] developed a “pseudo random walk” algorithm to find the lowest energy charge distributions. These studies showed that the basic sites are preferentially populated. Given that the total number of charges being considered here is always much less than the total number of basic sites, it was observed that the lowest energy distribution could always have charges located only on the basic sites.

In this method, contributions from Coulomb repulsion among protonated basic sites and relative protonation energies were included in calculating relative total energies. Using the crystal structure of cytochrome *c'* dimer, relative total potential energies were calculated using

$$\Delta E = \sum_{i=1}^{m-1} \sum_{j=i+1}^m \frac{q^2}{4\pi\epsilon_0 r_{ij}} + \sum_{i=1}^4 n_i \Delta E_{prot,i} , \quad (3.1)$$

in which ϵ_0 , r_{ij} , q , and m are the permittivity of free space, distance between two charged sites, charge on the protonated sites ($q = 1$ in this study), and total number of protonated sites. The first term accounts for the Coulomb repulsion from charge-charge interactions. The protonation energy of the basic amino acids, $\Delta E_{prot,i}$, is included in order to account for the protonation energy differences between basic sites. Here, n_i is the number of residues of type i , with $i = 1, \dots, 4$ representing arginine, histidine, lysine and the N-terminus (alanine), respectively. Protonation energies were calculated using *ab initio* methods with the Gaussian 03 program [93]. Geometry optimizations were performed with the MP2 method and the 6-311+G(d,p) basis set to obtain the total electronic energies of four protonated and non-protonated (neutral) basic amino acids in the gas phase. In order to calculate the lowest electronic energy for each basic amino acid, the most basic nitrogen atom in the side chain was protonated. The electronic energy difference between the protonated and non-protonated forms gave the protonation energies listed in Table 3.1.

The energy ordering of calculated protonation energies is also similar to the ordering of experimentally measured affinity and basicity values [94, 95]. Note that the electrostatic calculation of Eq. 3.1 is a crude one in that it ignores interactions between charged sites and the local protein environment. However, its use here is as a qualitative screening method only.

AminoAcid	Protonation Energy $\Delta E_{prot,i}/\text{kcal mol}^{-1}$
Arginine	-242.9
Histidine	-228.9
Lysine	-224.0
N-Terminus (Alanine)	-229.4

Table 3.1: Calculated gas-phase protonation energies of the basic amino acids using the MP2/6-311+G(d,p) method.

3.1.2 Results

Potential energies were first calculated by systematically constructing all possible permutations of charges among the basic sites, keeping a fixed charge partitioning between the two monomers. In other words, for the D10 charge state (with a total charge of +10 on the dimer) systems with fixed charge partitionings of M1/M9, M2/M8, M3/M7, M4/M6, and M5/M5 were formed (here Mx/My denotes x charges on one monomer and y charges on the other). For each charge partitioning, the potential energies of all possible charge locations consistent with the partitioning were evaluated using Eq. 3.1. The potential energies were ordered from lowest to highest, and were used to identify the most favorable charge distributions for each charge partitioning. This procedure was repeated with the D18 charge state.

The lowest potential energy found for each charge partitioning from this bare charge method is plotted in Figure 3.2 for the D10 and D18 total charge states. In both states, relative energies increase with charge asymmetry between the two monomers. These results demonstrate that symmetric charge distributions at both charge states (D10 and D18) are energetically favored at this level of modeling the electrostatic interactions. As well, the potential energy rises more steeply with charge asymmetry when the total charge is larger. The total potential energy can be divided into two contributions: the intermolecular energy representing the repulsion between the two monomers, and the intramolecular energy representing the repulsion among the charges within the same monomer. Intermolecular potential energies are also plotted as a function of charge partitioning for the D10 and D18 charge states in Figure 3.2. As the charge asymmetry grows, the intermolecular potential energy decreases while the intramolecular one increases. This of course simply reflects the fact that for asymmetric distributions, one monomer is preferentially charged, and hence incurs the greater electrostatic energy. As charge is concentrated more in a one monomer, the net repulsion between the two monomers in the dimer decreases. The key observation,

which will be expanded upon below, is that the barrier for dissociation of the protein complex is governed mostly by the intermolecular potential energy.

3.2 Empirical Potential Energy Calculations

3.2.1 MD Simulations

All MD simulations were carried out by using the Gromacs 3.2 software package [50–52]. Partial charges and force field parameters from the OPLS-AA/L force field [65, 68] were assigned to the cytochrome *c'* dimer. Protonated basic residual parameters are included in the published OPLS-AA/L force field but heme prosthetic group parameters are not. Therefore, in order to simulate cytochrome *c'*, we developed heme prosthetic group parameters and included them in the OPLS-AA/L force field.

P₄₅₀ heme parameters taken from the Gogonic group [96] were modified to match the heme group in cytochrome *c'*. The OPLS atom types for the heme prosthetic group are shown in Appendix B. Two oxygen atoms (ligand) included in P₄₅₀ were removed from the topology file. In order to neutralize the heme group, two hydrogen atoms were added to O2A and O1D oxygen atoms. Partial charges of the heme group to fit to cytochrome *c'* were taken from heme parameters developed for the Charmm force field by Autenrieth and co-workers [97]. Furthermore bond parameters were included between center atom (Fe) of heme and NE2 atom of histidine in order to describe the bond between heme and histidine residual in cytochrome *c'*. To represent the bond between Fe and NE2, all bond, bond angle, and dihedral parameters were obtained from Ref. 97. Dihedral parameters were modified to fit to the OPLS potential function. Modified dihedral parameter values are mostly estimated parameters.

For each total charge state and charge partitioning, charge distributions were chosen from the bare charge calculation for further studies using MD simulations. To achieve the specific charge state, charges were distributed (protonated) among selected basic sites (arginine, histidine, lysine, N-terminus) with +1 net charge in each residue distributed according to the OPLS-AA/L partial charge assignment. All other amino acid residues were kept neutral. Also, the heme group and the histidine residue bonded to the heme group were not protonated. Limited memory quasi-Newton method (L-BFGS) [98] energy minimizations were carried out to remove any bad contacts between atoms before all MD simulations. All covalent bond lengths were constrained to a 0.00001Å tolerance by the Shake algorithm [99]. Cutoff and periodic boundary conditions were not applied in these simulations because the system is an isolated protein dimer and no solvent is included. Initial velocities were generated according to a Maxwell-Boltzmann distribution at 300 K. The system temperature (300 K) was controlled by the Berendsen weak coupling scheme

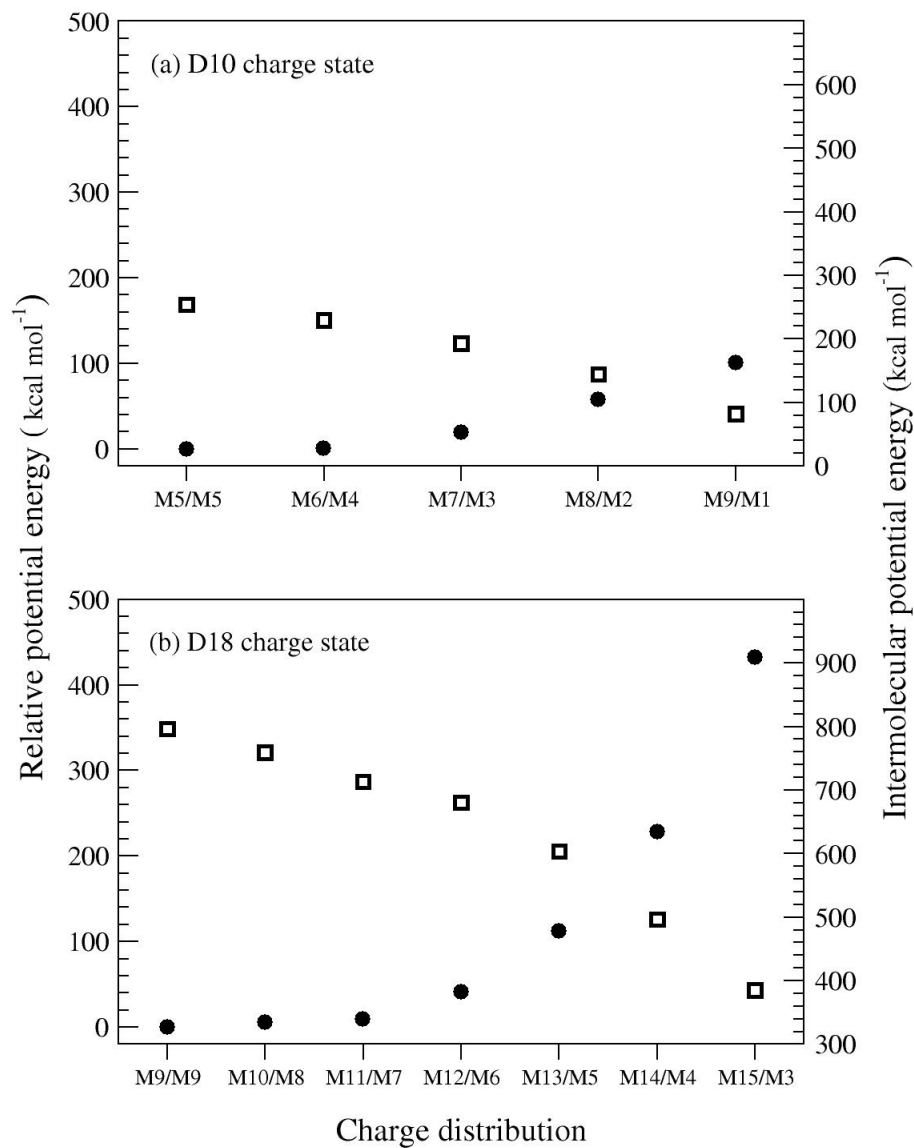


Figure 3.2: Relative potential energies (filled circles - read against the left vertical scale) and intermolecular potential energies (open squares - read against the right vertical scale) of the lowest energy charge configuration as a function of charge partitioning calculated using the bare charge method for the (a) D10 and (b) D18 total charge states. Here Mx/My denotes x charges on one monomer and y charges on the other.

with a relaxation time constant of 0.1 ps [72].

When employing semi-empirical potentials, the reference state for the zero of potential energy changes with system. Therefore, energies should be corrected according to the change in the number and location of protonated sites when different charge configurations are compared. When a neutral basic residue is protonated, the energy change should equal the protonation energy listed in Table 3.1. However, the OPLS-AA/L force field only partially accounts for this energy and the difference must be explicitly added. The potential energy correction term is thus

$$V_{corr} = \sum_{i=1}^4 n_i (\Delta E_{prot,i} - \Delta E_{MD,i}) , \quad (3.2)$$

in which $\Delta E_{MD,i}$ is the potential energy difference between the protonated and neutral amino acids after having performed L-BFGS energy minimizations of each with the OPLS-AA/L force field.

Vibrational frequencies of the bonds were not accounted for in this calculation because all bonds were kept frozen. All potential energies calculated from each MD simulation were corrected using

$$V = V_{MD} + V_{corr} , \quad (3.3)$$

in which V_{MD} is the total MD potential energy.

3.2.2 Results

In order to assess the utility of the bare charge method for selecting candidate charge configurations, two particular charge states were selected, M7/M3 and M10/M8. For each of these charge states, the energies of all possible charge configurations were calculated using the bare charge method, and formed into a gaussian like distribution. From each distribution, three groups were selected: (1) the 20,000 configurations with the lowest energies (denoted “low”), (2) 6000 configurations selected from the peak of the distribution with intermediate energies (denoted “medium”), and (3) the 6000 configurations with the highest energies (denoted “high”). Charge configurations with low energies were considered for further studies using MD simulation. Therefore, a large sample was selected for the lowest energy group while small samples were selected for the other two groups. Short simulations (1 ps) were performed for all the configurations in each group and average potential energies were calculated over the last 0.5 ps. These potential energy values were then formed into distributions, and are plotted in Figures 3.3 and 3.4.

In the bare charge calculation, the energies of the “low”, “medium”, and “high” configurations are separated from each other by more than 150 kcal/mol and have a very narrow spread. The distributions in Figures 3.3 and 3.4 show mean values that still maintain the low, medium, and high energy ordering but have total widths of more than approximately

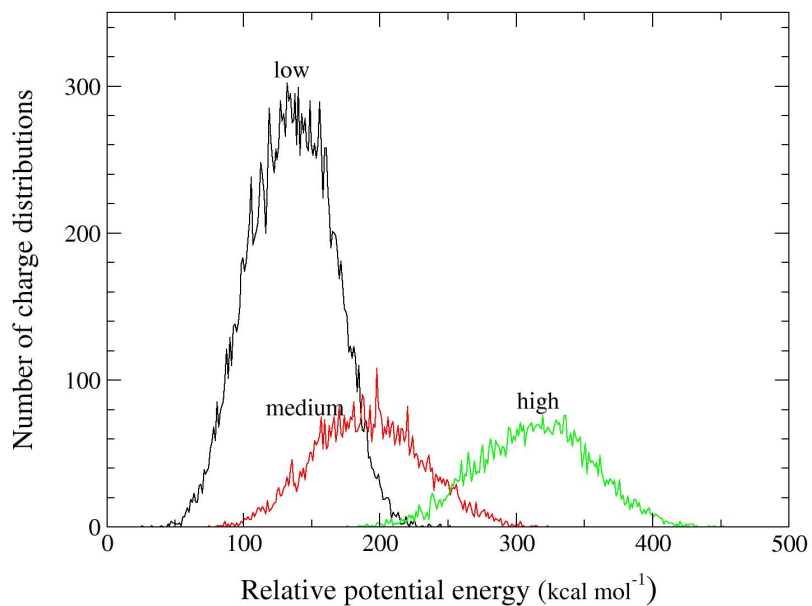


Figure 3.3: Distributions of relative potential energies for short run trajectories started with three groups of initial conditions, denoted “low”, “medium”, and “high”, taken from the bare charge calculation. These results are for the M7/M3 charge partitioning of the D10 charge state.

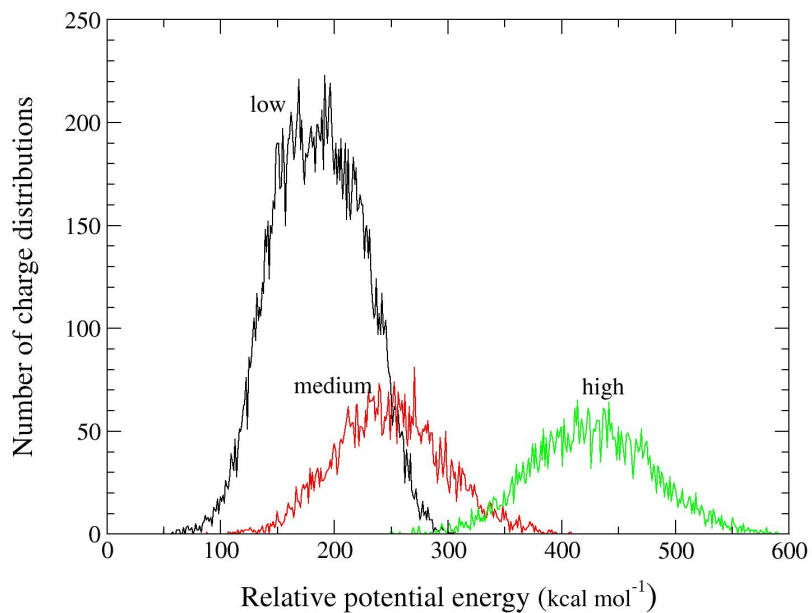


Figure 3.4: The same as Figure 3.3 except for the M10/M8 charge partitioning of the D18 charge state.

100 kcal/mol. There is substantial overlap among them, especially between the “low” and “medium” distributions. In other words, there are charge configurations that in the bare charge calculation had energies more than 100 kcal/mol above the minimum energy that relax in the short MD run to quite low energies. Stated another way, the configurations with the lowest energies in the bare charge calculation do not map directly to the lowest energies in the short MD calculation.

Two effects are included when moving to the short MD calculation. The first is that a better electrostatic model is used because the interactions are evaluated with the OPLS-AA/L semi-empirical potential. The second is that the MD run allows small adjustments in structure to remove any interactions that are particularly repulsive (but does not allow any larger structural changes to occur). The main conclusion to be drawn from Figures 3.3 and 3.4 is that no one single charge configuration will be important but rather distributions of configurations will contribute to the behavior of the system. The energies of these configurations are similar, and their relative ordering depends upon the particular instantaneous structure of the system. Thus, caution should be used when employing only a single structure, such as a crystal structure, to predict electrostatic energies in proteins when using an atomistic model.

For each of the D10 and D18 charge states and for each charge partitioning, the 20,000 charge configurations with the lowest energies, calculated with the bare charge method, were relaxed with short (1 ps) MD runs. From this MD set, the 50 lowest energy charge configurations were selected for longer MD runs. These simulations were conducted for up to 20 ps to better relax the system, and allow more accurate potential energy estimations to be made. Selecting only the single lowest energy charge configuration can lead to incorrect conclusions due to the limitation of sample size for short MD simulations. Fluctuations often cause the energy ordering of the configurations to change. Therefore, potential energies of the entire set of 50 charge configurations were averaged to estimate the lowest potential energy for a given charge partitioning.

The distribution of relative potential energies for different charge partitionings between two monomers in the D10 and the D18 charge states are shown in Figure 3.5. The energies, shown as filled squares, were calculated by averaging over time steps (energies were written every 1 fs) from 10 ps to 20 ps for each of the 50 charge configurations, and then averaging each of these resulting 50 energies. In addition to energy distributions from the 20 ps MD runs, the average energy distributions by averaging the same 50 configurations from 1 ps simulations are also shown as filled circles. The error bars indicate the widths of the distributions of the average energy of the ensembles of 50 trajectories. The average energies of individual trajectories are converged to statistical errors that are much less than the width of the distributions from the ensemble.

In addition to potential energy, the mean number of overcrossings for the two monomers

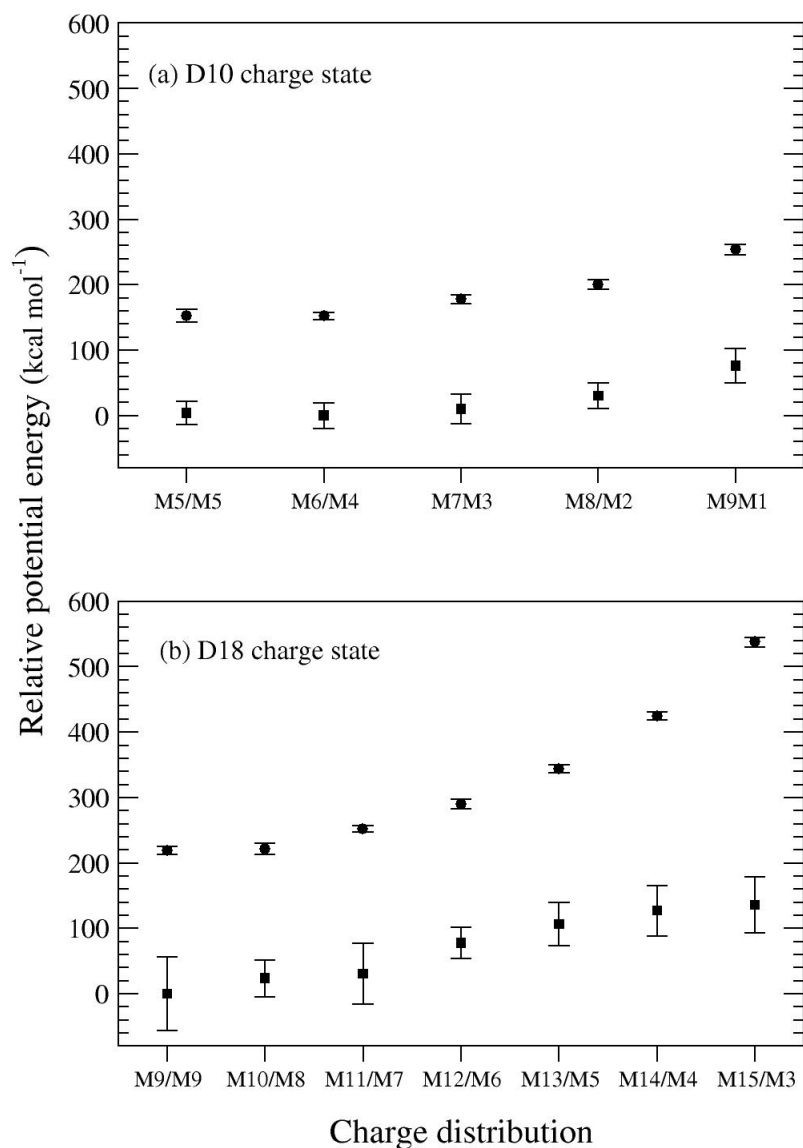


Figure 3.5: Averages of the relative potential energies at 1 ps (circles) and 20 ps (squares) of a MD run of an ensemble of 50 charge configurations selected with the lowest energy from the distributions in Figures 3.3 and 3.4, as a function of charge partitioning between two monomers in the (a) D10 and (b) D18 total charge states. Here M_x/M_y denotes x charges on one monomer and y charges on the other. The error bars indicate the widths of the distributions of the average energy of the ensemble.

was calculated every 0.1 ps time step during the 20 ps MD run and averaged over two time intervals (0.5 to 1.0 ps and 15.0 to 20.0 ps). Final values were then estimated by averaging each of 50 configurations for the two selected time intervals. Averages of the mean number of overcrossings for all charge partitionings for the D10 and D18 total charge states are listed in Table 3.2.

	1 ps		20 ps	
	chain 1*	chain 2*	chain 1*	chain 2*
D10 charge state				
M5/M5	71.0 ± 0.2	70.3 ± 0.4	70.9 ± 0.5	71.2 ± 0.5
M6/M4	70.7 ± 0.5	70.8 ± 0.4	71.0 ± 0.6	71.3 ± 0.6
M7/M3	71.0 ± 0.4	70.9 ± 0.3	70.7 ± 0.6	71.5 ± 0.6
M8/M2	70.8 ± 0.2	71.1 ± 0.3	70.1 ± 0.5	71.8 ± 0.5
M9/M1	70.4 ± 0.2	71.2 ± 0.4	69.7 ± 0.5	72.0 ± 0.4
D18 charge state				
M9/M9	70.0 ± 0.3	70.5 ± 0.4	68.4 ± 1.7	69.3 ± 2.5
M10/M8	70.1 ± 0.3	70.6 ± 0.3	66.8 ± 1.7	70.5 ± 1.0
M11/M7	70.0 ± 0.3	70.8 ± 0.4	66.5 ± 2.1	70.3 ± 1.5
M12/M6	69.7 ± 0.2	70.8 ± 0.4	66.1 ± 2.0	70.7 ± 0.9
M13/M5	69.4 ± 0.3	70.7 ± 0.4	64.7 ± 1.9	70.8 ± 0.8
M14/M4	69.0 ± 0.3	70.9 ± 0.4	60.9 ± 1.9	70.9 ± 0.7
M15/M3	69.0 ± 0.2	70.9 ± 0.3	58.0 ± 2.2	70.9 ± 0.7

Table 3.2: Average mean number of overcrossings taken at 1 ps and 20 ps for MD runs of a sample of 50 charge configurations selected with the lowest energy from the distributions in Figures 3.3 and 3.4. *Chains 1 and 2 denote the monomer with the higher and lower charges, respectively.

There are several observations that can be made from the data in Figure 3.5. First, for all cases, the potential energy is lowered by relaxations of the protein structure that can occur on the longer time scales. However, this relaxation is qualitatively different for the low (D10) and high (D18) total charge states.

For the D10 state, the energy lowering is uniform with charge partitioning so that the shape of the resulting curve is qualitatively the same both before and after relaxation. Examining the values in Table 3.2 shows that even after 20 ps, the average mean number

of overcrossings is almost unchanged from the values at short times. This indicates that no major structural changes have occurred. If one looks closely at the values, the values for chain 1 (the one with the higher charge) decrease slightly while those for chain 2 (the one with the lower charge) increase slightly on moving down a column for the 20 ps data. This means that the monomer with the higher charge state swells a little, and the one with the lower charge state contracts a little, as the degree of asymmetry increases.

The D18 state, on the other hand, shows qualitatively different behavior. The lowering of the relative potential energy is higher as the asymmetric partitioning increases, as seen in Figure 3.5. Furthermore, the average mean number of overcrossings, seen in the 20 ps data of Table 3.2, decrease markedly for the monomer with the higher charge but remain constant for the monomer of lower charge. This indicates that the monomer with the higher charge state undergoes partial unfolding. This unfolding increases the monomers surface area, and results in a lowering of the electrostatic repulsion of the charges located within it. This is the reason that the relative potential energy of the asymmetric partitioning is decreased more than that of the symmetric one by protein relaxation. The expanded monomer is not in a completely unfolded state, though. This can be seen in Figure 3.6 in which one trajectory from each of the ensembles was selected to give an indication of the structural changes that occur upon relaxation. The partial unfolding of the monomer with the greater charge in the D18 state is evident. Again, it must be emphasized that these structures are not unique, and that different trajectories in the ensemble relax to differing structures. Finally, it should also be noted that on the timescale of these calculations, many of the symmetric charge state species promptly dissociated (with little unfolding occurring during the dissociation event) while the asymmetric ones remained bound.

The second observation concerns the widths of the distributions of relative energies, indicated by the bars on the data points in Figure 3.5. Each individual trajectory has a well-converged average energy. The ensemble of 50 trajectories though has a distribution of energies, whose mean is plotted as filled circles, and whose widths are represented by the bars. Again, the energy ordering of the states changes upon protein relaxation. However, for the lower charge state, the band of energies for M5/M5, M6/M4, and M7/M3 overlap almost completely. This implies that there are manifolds of different charge partitioning and configurations that, upon protein relaxation, have similar energies. In other words, for low charge states, the monomers act as sponges, able to absorb a certain number of charges without any noticeable change in structure or energy. This implies that there is no single charge partitioning or configuration that dominates the behavior of the system, and one should not extrapolate results from any single structure (such as a crystal structure). Also, at low charge states, a fair degree of asymmetry can be energetically stabilized, even in the absence of any significant change in the structure of the complex.

For the higher charge state though, this breaks down due both to the extra Coulomb

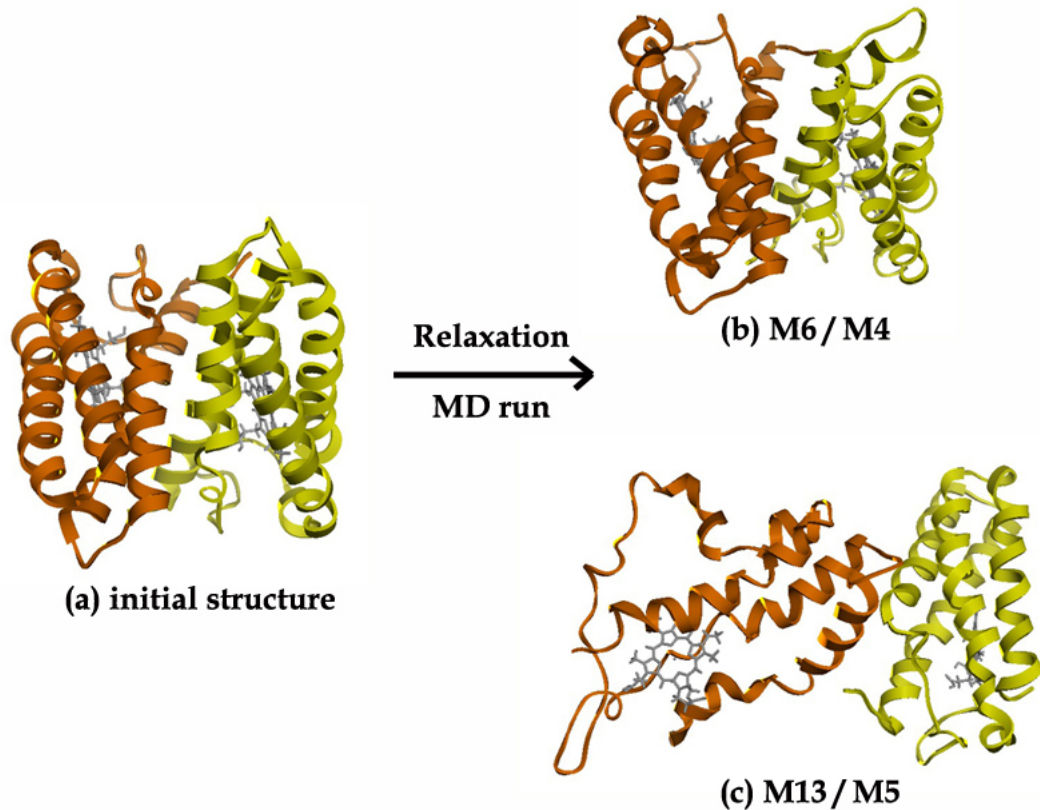


Figure 3.6: Snapshots of cytochrome c' structures initially (a) and after a 20 ps MD simulation with the (b) M6/M4 and (c) M13/M5 charge partitioning. In each complex, the higher and lower charged monomers are drawn to the left and right respectively.

repulsion present in the system, and to the decreased flexibility in the choice of sites to locate charges, so that even after accounting for the widths of the distributions, the symmetric distribution is still energetically favored.

3.3 Discussion

3.3.1 Coulomb Repulsion Model

The MD results presented here use more realistic electrostatic models than had been used in a previous study of ellipsoids [20] but point to essentially the same conclusions, namely that the system is governed predominantly by electrostatic repulsions between the net charges. This, of course, is not a novel statement and has been made by several researchers examining

this phenomena. For example, Kaltashov *et al.* [100] demonstrated a correlation between the total charge and the solvent accessible surface area of proteins in the ESI-MS process. However, it is worth enumerating the consequences of this model because the full extent of it has not been employed.

The Coulomb repulsion model predicts two general trends.

- A.** Charges that are free to move adopt the lowest energy when uniformly spread over the surface of an object, with a slight concentration among the points that are furthest apart if polarization effects are accounted for. To remain at lowest energy, the surface charge density should be kept constant.
- B.** For an object, such as a protein complex ion that is a collection of like objects bound together, all other interactions being equal, to the extent that Coulomb intermolecular repulsion dominates, the lowest energy barrier for dissociation occurs when all fragment ions have the same charge, that is the total charge is divided symmetrically among the fragment ions.

The second trend follows simply from the fact that the long-range repulsion among charges is greatest for the symmetric case. For example, for a complex that dissociates into two fragments with charges q_1 and q_2 , the long-range repulsion, $q_1 q_2 / 4\pi\epsilon_0 r$, is a maximum when $q_1 = q_2 = n/2$ (if $q_1 + q_2 = n$), and progressively decreases as the difference between q_1 and q_2 increases. As the charge asymmetry increases, so does the barrier to dissociation. This is also consistent with the experimental results of Sinelnikov *et al.* [18] who assert that the long-range repulsion between charged groups dominate the dissociation of a number of protein complexes, and who sought to understand the relationship between measured activation energies and some calculated values of this repulsion.

It should also be noted that this long-range repulsion is generally the same regardless of how the charges q_1 and q_2 are distributed, since the repulsion between two uniformly charged spheres with charges q_1 and q_2 is the same as that between two point charges. In other words, the barrier to dissociation is determined by the absolute total charge, not on the surface charge density.

When considering protein complexes, it is important to stress that the minimum electrostatic state in Trend A is never reached in practice because the charges in proteins are mostly localized at particular sites, so are unable to spread into a layer of uniform charge density. Thus, **all** charge configurations in protein complexes have energies higher than this global minimum, so in principle can have their energies changed relative to one another by relaxation processes. In other words, a charge distribution that is slightly non-uniform might be relaxed enough by protein structural changes to have the same energy as one that is more uniform. This is the reason that states of differing charge configuration and

partitioning, as seen in Figure 3.5, can exist at approximately the same energy. Thus, in protein complexes there will always be fluctuations around the constant charge density limit of Trend A. This result is consistent with the calculations using charged ellipsoids [20], in which case fluctuations of 10-15% were observed from the constant surface charge density state.

Finally, Trend A depends upon the mutual repulsion of charges within a complex, which in turn depends upon the surface charge density. The greater the surface charge density, the greater will be this Coulomb energy relative to other interactions in the complex, and the more Trend A should dominate. On the other hand, Trend B should dominate as the total charge on the complex increases. In typical protein complexes, the binding energy holding a given monomer in a complex does not scale with the size of the complex but rather depends only upon a constant number of binding interactions. However, the charge that can be imparted to such a complex grows with its size. Thus, the repulsion between the charges can quickly dominate as a complex grows larger. Note that this is different from the underlying principles involved with charged droplet or fission models. In those cases, Coulomb repulsion is balanced by surface tension, the latter of which does scale with surface area. The balance between these two forces then determines the behavior of the system. Finally, in principle, the surface charge density and total charge can be varied independently.

3.3.2 Application of the Coulomb Repulsion Model

Consider for a moment the application of the Coulomb repulsion model to large, multimeric complexes. In many cases, large multimeric complexes are produced with high charges but small charge per monomer ratios, that is small surface charge densities. In this case, one expects Trend B to dominate while there may be fluctuations around the limit suggested by Trend A. In the equilibrium state, a complex composed of the same monomers with a total of n charges should exist in a state in which the charges are generally evenly distributed among the complex, as suggested by Trend A. However, as seen in Figure 3.5, because the surface charge density is typically low, it is expected that a range of charge partitionings will be present, with some monomers in the complex, at any given time, having a slight concentration of charge, but undergoing small relaxations to stabilize them. That is, the complex ions should be composed of a range of charge configurations and partitionings close to evenly distributed but also with some asymmetry.

When energy is imparted to such a complex, through SORI-CAD, BIRD, or other gentle heating techniques, the internal energy slowly increases. Those complexes that in the equilibrium state have one monomer with a slight charging, are expected to be affected first since the greater intramolecular repulsion present in these monomers will make them

more susceptible to conformational changes. These conformational changes can eventually cause these monomers to be ejected from the complex, much like the strength of a chain is determined by its weakest link. This is postulated to be the main reason experimental studies show that multimeric complexes tend to dissociate almost exclusively by losing only a single monomer at a time.

As the internal energy increases, a complex has many dissociation channels available for different charge partitionings and configurations. If the energy increase is slow relative to the timescales of protein structural changes, the lowest energy barrier will be, according to Trend B, the one that divides the charge evenly. However, in order for this dissociation pathway to be accessible, in accordance with Trend A, the surface areas of the two dividing fragments must be approximately the same. If a monomer divides from a multimeric complex, it must partially or completely unfold up to the point where its surface area equals that of the other fragment ion in the dissociation channel. The proposal that monomer unfolding is inherent in the dissociative pathway has been put forth by several research groups [6, 12] and examined in particular detail by Jurchen and Williams [19]. We employ this idea as a means of explaining how surface charge density can be kept constant during the dissociation process. If, either due to structural constraints within the monomer that limit unfolding, or due to a size disparity between the remaining fragment ion and the fully unfolded monomer, the surface areas of the monomer and remaining fragment ion cannot be made the same, then the symmetric charge dissociation pathway will be inaccessible.

In this case, the system tries to get as close to the symmetric limit as possible by putting as much charge on the departing monomer as allowed by its surface area, according to Trend A. Thus, the ratio of the surface area of the unfolded ejected monomer to that of the remaining complex should give a reasonably good approximation of the charge ratio between the two.

These scenarios have been observed experimentally. For example, tetrameric complexes of protonated streptavidin ions are found to fragment in a symmetric fashion [6, 18], with S_4^{+n} ions producing predominantly $S_3^{+n/2}$ and $S^{+n/2}$ fragments, with n varying between 14 and 18. Fluctuations around these limits of one unit of charge are also observed. This indicates that one streptavidin monomer is sufficiently flexible so as to unfold to produce a surface area commensurate with the remaining trimer fragment ion. This interpretation is consistent with the electrostatic model calculations of Sinelnikov *et al.* [18] who showed that a partially unfolded monomer subunit could account for the observed charge ratios. Interpreted in another way, the monomer need unfold only to the extent of making its surface area commensurate with the remaining trimer fragment ion.

Similar results have been measured for the dissociation of the tetrameric complex transthyretin (TTR) by Sobott *et al.* [13]. They dissociate complex ions with +15 charge, and observe that the dominate dissociation pathway produces monomer ions of charge +8 and

trimer ions of charge +7. In addition, monomer ions of charge +9 and +6 are also observed. Thus, in this case symmetric charge dissociation is observed, consistent with predictions of the Coulomb repulsion model.

For pentameric complexes of the Shiga toxins, the results of Klassen and coworkers [12, 18, 101] are slightly different. The +14 charge state complex divides into a monomer and tetramer with a broad range of partitionings around a dominant channel which has a monomer being ejected with charge +6. The symmetric channel has very low population. In this case, the Coulomb repulsion model would rationalize that the complex has not been able to reach the symmetric limit because the monomer has not been able to unfold to a state with surface area large enough to match the tetrameric fragment. The dissociation is “near symmetric” though, in that the division of charge is close to the symmetric limit. The electrostatic calculations of Sinelnikov *et al.* [18] support this interpretation. In that work, a partially unfolded structure is found to reproduce the experimental charge partitioning, and this is physically attributed to a disulfide bond in the monomers which prevent complete unfolding. However, in addition, they estimated the charge partitioning in the limit of complete unfolding by breaking this bond, in their fully unfolded protein structure model (FU-PSM). The reported values, shown in the last column of Table 2 in Ref. 18, are almost precisely half the total charge state of the original Shiga toxin complexes. In other words, if the monomer were able to completely unfold, the symmetric dissociation channel would be the dominant one.

The measurements of Sinelnikov *et al.* [18] are particularly interesting because they examine both positive and negative total charge states of the Shiga toxin complexes. Generally, they find that the charge fragmentation channels are the same, regardless of the sign of the charge on the complex. Again, this is consistent with the Coulomb repulsion model, since the long-range dissociation repulsion depends only upon the product of the charges on the fragment ions, and not on their absolute signs. Trend B would predict the same dissociation barrier for complexes with the same total charge but differing signs.

Larger multimeric complexes behave in a similar way. Robinson and coworkers [14, 17] have studied 12-mers of *Ta*HSP16.9 and 24-mers of *Mj*HSP16.5 at high charge states (approximately +2 charges per monomer). Again, because the surface charge density is low, it is expected that fluctuations around uniform charge partitioning will be present in the unperturbed ions, and that some complexes will contain monomers that are charged higher than the average. These are the ones that are expected to be most susceptible to dissociation when energy is added to the system.

The large charges on these complexes should make Trend B dominate. Thus, the complexes should try to dissociate into fragments of equal charge and surface area. However, because the complexes are so large, even completely unfolded monomers will not have surface areas equal to the remaining complex fragment ions. In this situation, the model

predicts the charge partitioning between monomer and complex fragment ions to equal the ratio of surface areas of the two. This prediction is consistent with reported surface area ratio calculations [17]. In fact, it was observed experimentally that an n -mer complex ion can lose a single monomer ion, with the resulting $(n-1)$ -mer fragment ion again dissociating by losing a single monomer ion to produce a lower charged $(n-2)$ -mer complex ion. In each case, the ratio of surface areas of the complex ion and unfolded ejected monomer ion approximated the charge partitioning among the fragments. This behavior is also consistent with the Coulomb repulsion model. After ejecting a single monomer, a complex ion may still have a substantial charge, and thus still be dominated by electrostatic interactions. One then expects precisely the same behavior as with the original complex, namely additional loss of a monomer with the system attempting to attain as close to symmetric charge distribution as possible. This process is expected to repeat itself until the total charge of the remaining fragment complex ion is low enough that other interactions begin to compete with Coulomb repulsion.

Heck and coworkers [43] reported interesting dissociation behavior for the dissociation of tetrameric complexes of 2-keto-3-deoxyarabinonate dehydratase. These complexes dissociate almost exclusively into two dimeric fragment ions rather than the usual loss of a single monomer. The complexes are composed of dimers that have more binding interactions internally than between them, making them more prone to dissociate into dimeric fragments. They found in fact that the complexes are quite susceptible to dissociation. The key observation though is that the dimers are produced with near symmetric charges. This is completely consistent with Trend B which predicts that symmetric charge partitioning should dominate regardless of how a complex breaks into fragments.

It should be emphasized that charge-to-mass ratio is not a relevant parameter in determining the charge partitioning of multimeric ions. The relevant parameters are total charge on the ion, and surface charge density. So, for example, the dissociation of the tetrameric streptavidin ion into monomer and trimer ions of equal charge has been classified as an asymmetric dissociation because the charge to mass ratios of the resulting fragments are very different. However, within the Coulomb repulsion model, this dissociation would in fact be classified as symmetric, because the charge division is equal.

All the arguments above hinge upon two important assumptions: 1) that charge transfer is labile (and that sufficient charge sites are available), and 2) that monomer conformational changes occur on a timescale that is faster than the dissociation timescale. A number of experimental measurements [12, 18, 102–105] indicate that indeed charge transfer in proteins is labile. Some theoretical calculations [106–108] also indicate that proton migration along the backbone may be a likely mechanism for charge transfer. However, changing experimental conditions in such a way as to weaken these two assumptions will also weaken the general predictions of the Coulomb repulsion model.

For example, Felitsyn *et al.* [45] have reported measurements of the dissociation of ecotin dimer ions with high charges (+14 to +17). One would expect the Coulomb repulsion model to be applicable, and that symmetric dissociation is expected. However, while the symmetric dissociation channel was indeed observed, there was also significant population in channels that deviated from symmetry by one and two units of charge. The time dependence of the fragment distributions indicated that evidence for charge transfer was not strong. It was concluded that the different dissociation channels reflected the distribution of charge partitioning among the monomers in the original complex ions. This is completely consistent with the phenomena shown in Fig. 4 in which fluctuations about the symmetric distribution are expected simply due to local relaxation effects. If these dimers dissociate without charge transfer occurring, then one would expect to reveal the distribution of charge partitioning present in the equilibrium state which is distributed about the symmetric one. Note that during the ESI process, charges are expected to be mobile before the naked ion is produced. Thus, the equilibrium state is expected to be established in the complex, even if charge transfer is inhibited afterwards.

This is also seen in the surface-induced dissociation (SID) experiments of Wysocki and coworkers [9] in which ions are dissociated into fragments on a picosecond timescale by colliding them with a surface. This causes fragmentation to occur on a timescale that is faster than that for conformational changes in the monomers. For SID of cytochrome c dimers of charge +11, monomer fragments with charges +5 and +6 dominate. That is, a symmetric charge distribution is observed. In general, if experimental conditions are used that dissociate ions quickly compared with the timescale for monomer conformational changes, the Coulomb repulsion model predicts that each monomer in a complex should have approximately the same charge. Thus, monomers ejected from an n -mer would be expected to carry a charge of approximately $1/n$ of the total charge on the complex. In such cases for $n \neq 2$, the dissociation channel would be labeled as “asymmetric” within the Coulomb repulsion model because the fragment ions have differing charges, even though the charge to mass ratios of the fragment ions would be the same.

The Coulomb repulsion model is also consistent with the detailed experiments performed by Jurchen and Williams [19] that showed, among other things, the relationship between monomer flexibility and the degree of asymmetry in fragment ions. In particular, monomers with greater rigidity led preferentially to symmetrically charged dissociation products. Considering Trends A and B, rigid monomers are not able to change their surface areas much, due to conformational constraints.

Note that in all cases, one would never expect to see a multimeric complex ion dissociate by ejecting a monomer with greater than half the total charge on the complex (apart from small fluctuations).

Deviations from symmetric charge dissociation can be expected for ions with lower

charge states because other interactions compete with the Coulomb repulsion among the net charges. In this case, a more complicated dissociation mechanism is at work. It may happen that even the symmetric distribution of charge produces a barrier that is high relative to the internal energy, and the system needs to find another pathway for dissociation.

Asymmetric charge distributions lower the intermolecular repulsion but increase the intramolecular repulsion. In order for dissociation to occur, typically hydrogen bond interactions in the protein complex must be broken. Greater intramolecular repulsion can promote conformation changes in a monomer that increase its size, and thereby may weaken hydrogen bonds that bind the complex. In this way, the lowest energy path for dissociation can be through unfolding-like transition states encouraged by asymmetric charge distributions, as has been postulated by several groups [6, 12, 18, 33].

Chapter 4

Dissociation Barrier Estimation*

Free energy profiles were calculated to gain insight into the dissociation mechanism. More specifically, constrained molecular dynamics calculations were used to estimate the free energy changes as a function of the distance between the centers of mass of two monomers in a cytochrome *c*' dimer. A number of different charge partitionings were examined, as well as the behavior of the neutral complex. In addition to the relative free energies, structural changes were monitored by calculating the mean number of overcrossings and radius of gyration.

4.1 Method

Results from Chapter 3 show that the symmetric charge partitioning between two fragments is always favorable for both lower and higher charge states. These observations should also be applicable to the dissociation of homodimers. Experimental results show symmetric charge partitioning during the dissociation of homodimers with large total charge [19]. However, several research groups have observed a predominant asymmetric dissociation pattern in homodimers with small total charges under some conditions [9, 19, 33, 45]. Furthermore, the Coulomb repulsion model proposed in Section 3.3.1 is less applicable for explaining the dissociation of complexes with low charges. In this case, the interactions within a complex compete with the repulsion between its charges. Thus, a lower charged state was selected for further studies.

The D10 charge state (with a total charge of +10 on the dimer) with fixed charge partitionings of M1/M9, M2/M8, M3/M7, M4/M6, and M5/M5 was selected to study the dissociation mechanism in more detail. Ultimately, for each charge partitioning, 10 configurations were chosen from the MD simulations of Chapter 3. Here, the term “charge partitioning” refers to the number of charges that are assigned to each monomer in a complex ion. The term “charge configuration” refers to the particular arrangement of charges on charge sites. The 5 configurations with the lowest average energy from the 20 ps MD simulations were first selected. These were then augmented with an additional 5 configura-

*Results presented in this chapter appear in Wanasundara, S. N.; Thachuk, M., Free energy barrier estimation for the dissociation of charged protein complexes in the gas phase, *J. Phys. Chem. A*, 2009, 113, 3814-3821.

tions of lowest average energy chosen from the 1 ps MD simulations, ensuring that these did not duplicate any of the first five. It is important to sample different charge configurations for each charge partitioning because we expect several to contribute to experimental observations. Ultimately, this procedure produced 10 unique charge configurations of lowest energy for each particular charge partitioning. For each charge configuration, 5 trajectories were run, using different initial velocities, so that in the end, all reported averages included 50 different trajectories that sampled both charge configuration space as well as the phase space associated with a given charge configuration.

Each of the trajectories was initialized by starting with the crystal structure and placing charges in the appropriate locations for each configuration. The resulting structures were first minimized with the L-BFGS method and then equilibrated with 100 ps MD runs with the same MD conditions mentioned in Section 3.2.1 before starting COM constraint MD simulations.

The COM distance can be constrained during a MD simulation by using the Shake algorithm, and calculating the resulting constraint force. One has the option of keeping the COM fixed at precisely a given value, or of having the COM distance change as a function of time, so that the constraint moves as the simulation progresses. Ultimately, average constraint forces as a function of center of mass(COM) distance are being sought. It should be noted that an error in the pull code of Gromacs 3.2 version was corrected and details of the correction are given in Appendix D.

Our first attempts began by starting trajectories from the equilibrated bound dimer structure and then performing constraint MD simulations that changed the COM distance as a function of time from small to large values, spanning configurations from the bound dimer to its dissociated monomers. These simulations ran for approximately 100 ps so were quite rapid compared with the timescales required for structural changes in the complex. From each of these trajectories, initial positions and velocities were selected at a grid of COM distances, and these were used as initial conditions for new simulations, this time keeping the COM distance fixed at the grid values. We found that this procedure generated results that were difficult to converge and gave constraint forces that varied greatly at different COM distances. The problem appeared to be with the initial 100 ps trajectories that scanned the COM distances. The scanning rate was so fast that the complex could not adjust to changes in the COM distance, and hence structures were produced far from their relaxed states. In principle, if enough of these rapid scan trajectories could be run, meaningful statistics could be built up over time. However, the computational effort involved with doing so was too great.

Instead, we opted for a different method that involved slow changes in COM distance. In this method, the COM distance is changed by a small amount using a short MD run. This is followed by an equilibration MD run with the COM fixed at the new distance.

The trajectory at the end of the equilibration run is then used to initialize another short simulation during which the COM distance is changed again by a small amount. This is again followed by an equilibration run and the procedure is repeated many times until the range of COM distances is covered. In this method, each trajectory is initialized from the bound state equilibrated dimer structure and is systematically propagated through all the COM distances.

More specifically, the COM distance was changed by 1 Å at a time during short MD runs lasting 10 ps, so that the COM distance was changed at a rate of 0.1 Å/ps. The subsequent equilibration run lasted 100 ps before the next change of COM distance. Starting from the bound dimer state, the COM distance was both decreased to lower values and increased to large values. In this way, a broad range of COM distances was covered, and at each distance, constraint forces were calculated from the 100 ps equilibration runs. We found that the resulting forces were much smoother functions of COM distance and also converged more rapidly. We attributed this to the fact that the equilibration runs at each distance allowed the complex to relax at the new COM distance, and this relaxed structure provided the initial conditions for the next COM distance calculation. In this way, the complex has time to adjust to the changing COM distance, so that the method is closer to a slow growth process compared to one that initializes trajectories that are not equilibrated.

4.2 Results

Distributions of average constraint forces for different charge partitionings are shown in Figure 4.1. Here, the constraint force represents the intermolecular attractive or repulsive force caused by interactions between the monomers. The attractive forces are reported as negative numbers while the repulsive forces are reported as positive numbers. The ensemble averages of the constraint forces were calculated by averaging over time steps (constraint forces were written every 1 fs) from 50 to 100 ps in each of the fixed COM distance constraint MD runs. Constraint forces from the five different trajectories of a given charge configuration were averaged, and then the resulting forces from each of the 10 charge configurations were averaged. Error bars in Figure 4.1 indicate the standard deviation of the constraint forces calculated from each charge configuration. It should be noted that for any given trajectory the average constraint force is well converged. The fluctuations seen in Figure 4.1 show the variation of these forces as calculated with different trajectories. These fluctuations are likely due to variations caused by differing charge configurations, and by incomplete equilibration for trajectories within the simulated time period.

It is necessary to integrate the data in Figure 4.1 to obtain an estimate of the work, W , performed on the system at each fixed COM distance. This was calculated for each trajectory by multiplying the constraint force by the change in COM distance. The values

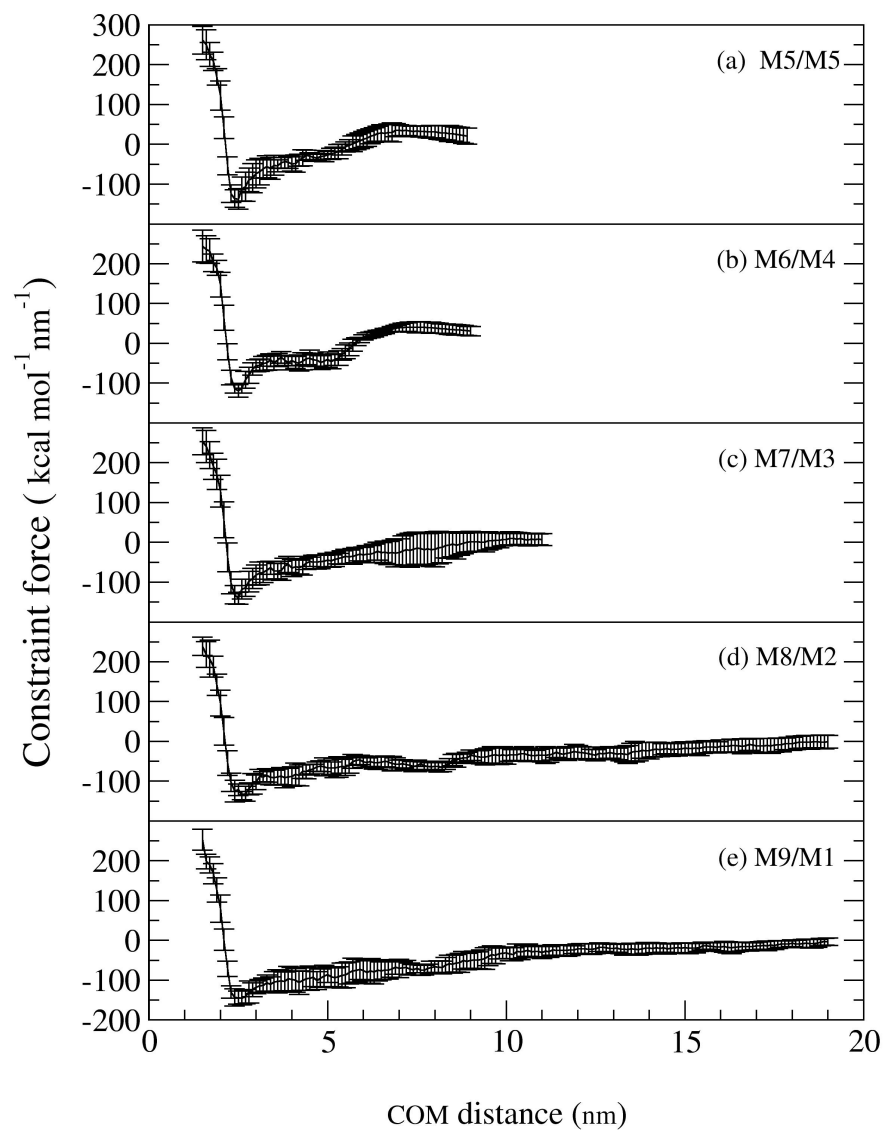


Figure 4.1: Constraint force as a function of center of mass distance between two monomers for the (a) M5/M5 (b) M6/M4 (c) M7/M3 (d) M8/M2 and (e) M9/M1 charge partitions. These are each averages over 50 different trajectories. The errors bar indicate the standard deviation of the average values.

of the 5 trajectories of a given charge configuration were averaged. This then produced 10 different values of work, for each of the 10 different charge configurations. The average work \bar{W} was calculated by averaging these 10 values, and the value of σ was then determined. The dissipative work was subtracted from the average work in order to approximate the free energy difference (Eq. 2.35) between two fixed COM distances.

Average free energy changes as a function of COM distance are plotted for different charge partitionings and for the neutral dimer in Figure 4.2. The minimum relative free energy for each curve was set to zero at the equilibrium distance of the dimer. Significant bound states are observed in each case, indicating that the dimer is stable for all the charge partitionings considered. All free energy minima for the bound states exist around 2 nm COM distance. Furthermore, all the free energy curves are essentially identical up to a COM distance of 2.8 nm, regardless of charge partitioning. However, once past this distance, the curves start to diverge according to the charge partitioning.

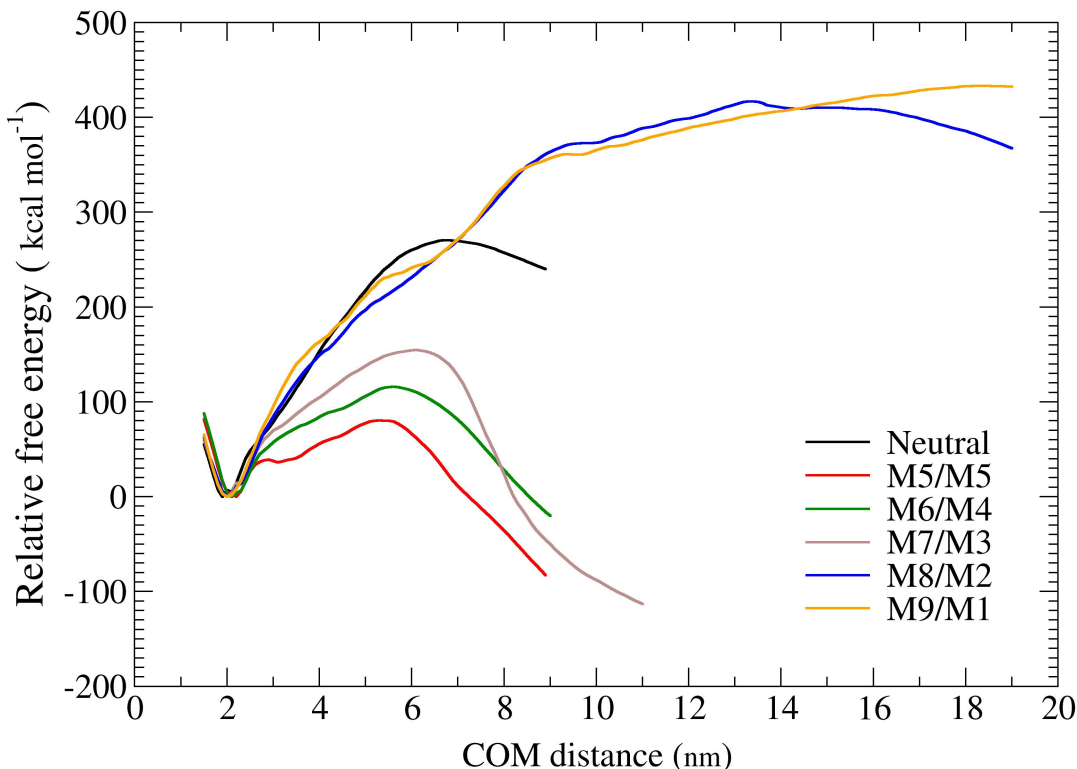


Figure 4.2: Relative free energy changes as a function of COM distance for different charge partitionings. Here M_x/M_y denotes x charges on one monomer and y charges on the other.

Energy barriers are present for the neutral complex, as well as the M5/M5, M6/M4, and M7/M3 charge partitionings within the simulated COM distance. The qualitative magnitudes of the barriers follow the pattern expected from the Coulomb repulsion model (Section 3.3.1) namely, that the lowest barrier is for the symmetrically charged complex (M5/M5) with the remaining ones increasing as the degree of charge asymmetry increases. These barriers are all below that for the neutral complex. This trend simply reflects the increase in intermolecular repulsion (which varies roughly as the product of the monomer charges) as the charge partitioning becomes more symmetric. The Coulomb repulsion from these charges reduces the barrier to dissociation, as compared with the neutral complex.

The curves for the M8/M2 and M9/M1 charge partitionings begin with the expected behavior at small COM distances. Between 2 and 6 nm, the M8/M2 is below the M9/M1 curve, and both are slightly below that for the neutral complex. This is the expected trend based upon the much lower intermolecular repulsion present with these very asymmetric charge partitionings. However, at a COM distance of approximately 6 nm, the M8/M2 and M9/M1 curves begin to increase substantially, eventually surpassing that even for the neutral complex. The top of a barrier is barely discernible for the M8/M2 partitioning, and for the M9/M1 one, it appears as if the relative free energy is still increasing, even at 19 nm. To help show the cause of this behavior, a number of properties will be presented.

The mean number of overcrossings for each of the two monomers in the dimer was calculated every 0.1 ps during the 100 ps COM constraint MD runs and averaged over the simulation time period. The values for the 5 trajectories of each charge partitioning were first averaged, and then the values were averaged across the set of 10 different charge partitionings. Averages of the mean number of overcrossings for all charge partitionings are plotted in Figure 4.3 as a function of COM distance. In addition, average radius of gyration values were also calculated and plotted in Figure 4.4.

The mean number of overcrossings for both monomers in the M5/M5, M6/M4, and M7/M3 charge partitionings remains essentially unchanged from small COM distances to those near the top of the free energy barrier. The average radius of gyration also stays constant over the same distances. This implies that no major structural changes occur. However, at larger distances, the number of overcrossings for both monomers start to decrease, with the monomer of higher charge decreasing the most. This decrease signals the occurrence of a structural change. Further, the increase of radius of gyration confirms the structural changes of both monomers at this stage. Figure 4.5b shows snapshots of the complex for particular trajectories at different COM distances. At a distance of 6 nm, the M5/M5 complex is clearly dissociated, and the monomers have structures that are similar to the ground-state ones, although they are somewhat relaxed. This relaxation causes the mean number of overcrossings to decrease. The monomers with the greater charge have greater intramolecular repulsion and thus relax to a greater extent.

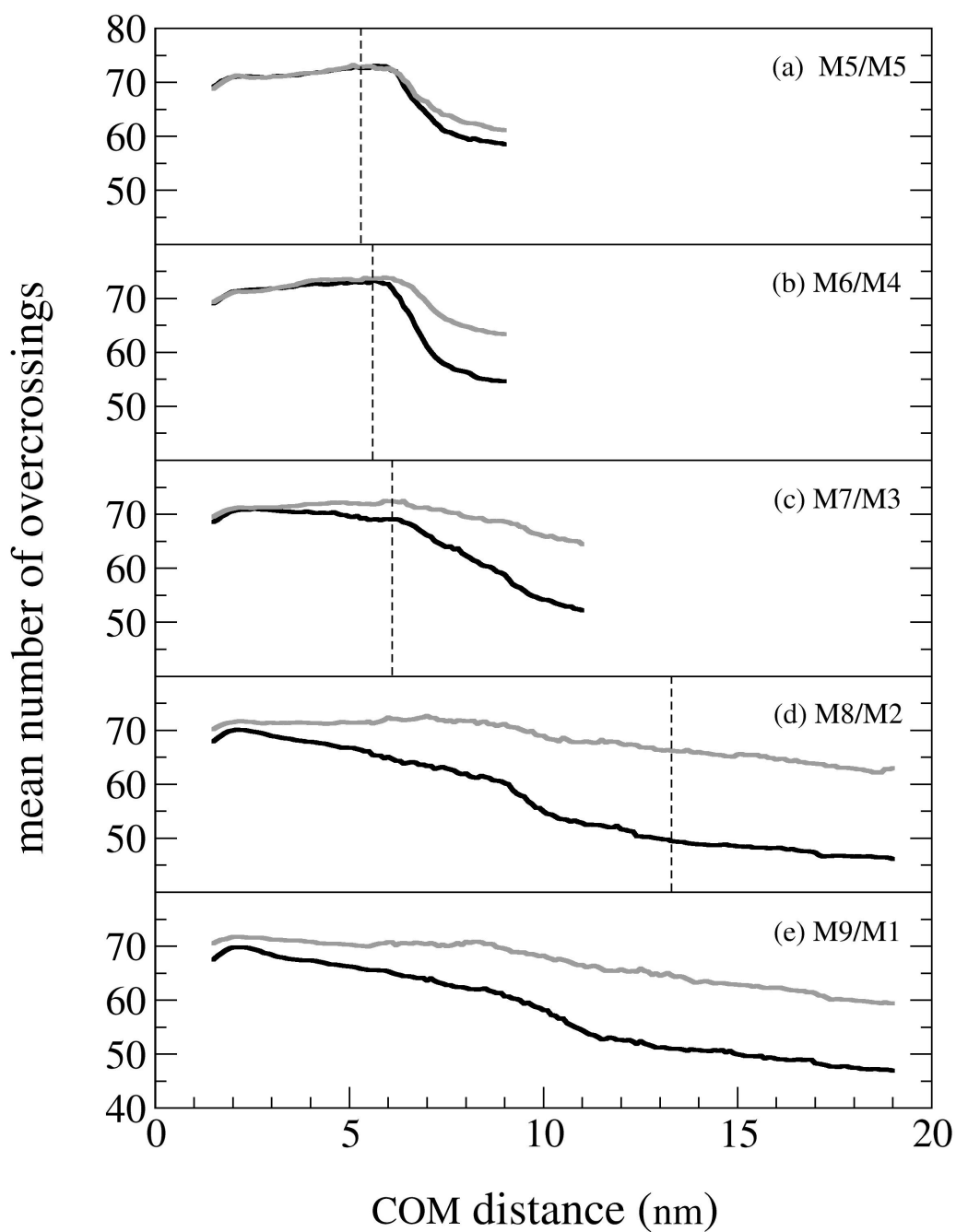


Figure 4.3: Average mean number of overcrossings as a function of COM distance for the (a) M5/M5 (b) M6/M4 (c) M7/M3 (d) M8/M2 and (e) M9/M1 charge partitionings. Vertical bars indicate the positions of the barriers in Figure 4.2. In each panel, the upper and lower curves correspond to values for the monomer with the smaller and larger charges, respectively.

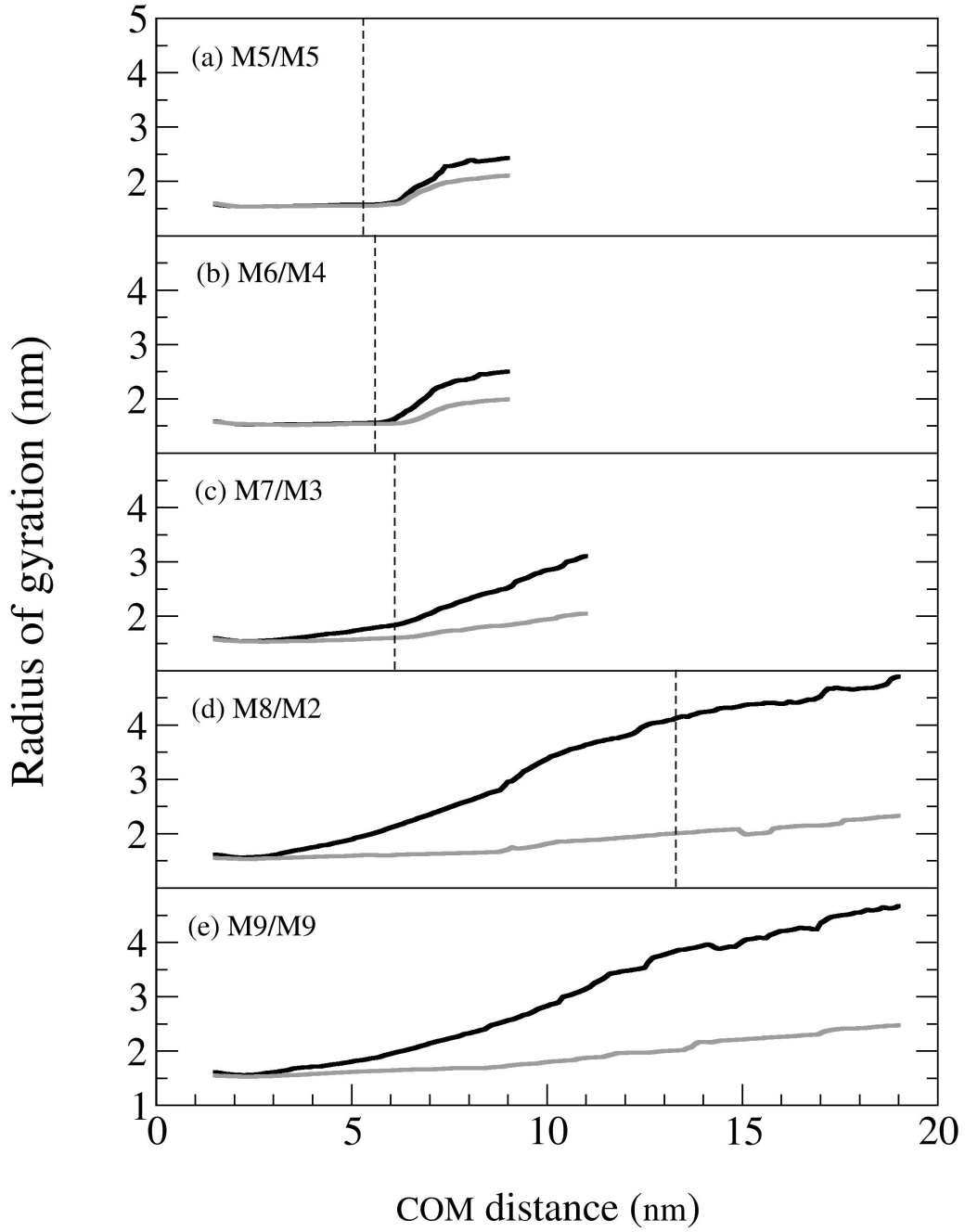
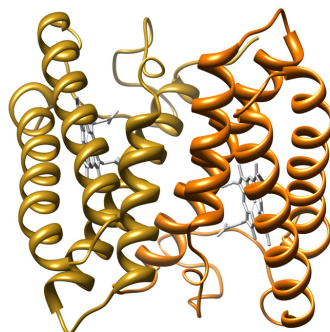
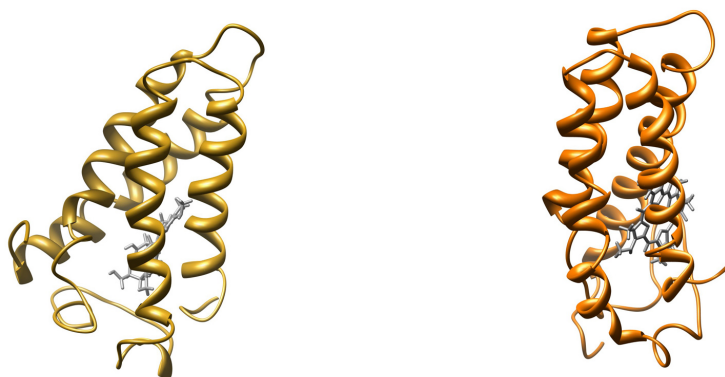


Figure 4.4: Average radius of gyration as a function of COM distance for the (a) M5/M5 (b) M6/M4 (c) M7/M3 (d) M8/M2 and (e) M9/M1 charge partitionings. Vertical bars indicate the positions of the barriers in Fig. 4.2. In each panel, the lower and upper curves correspond to values for the monomer with the smaller and larger charges, respectively.

(a)



(b)



(c)



Figure 4.5: Snapshots from the constraint MD run at a COM distance of 6 nm for the (a) M5/M5 and (b) M8/M2 charge partitionings (the monomer with the higher charge is shown on the left).

Consider now the behavior of the M8/M2 and M9/M1 charge partitionings in Figures 4.3 and 4.4. For these cases, as the COM distance increases from the bound state value, the mean number of overcrossings starts to decrease while the radius of gyration starts to increase. In these cases, structural changes begin to occur even before the barrier in the free energy curve is reached. Again, the snapshot in Figure 4.5c offers a representative view of the types of structural changes that occur. At larger distances, the monomer with the larger charge is significantly unfolded, with the helical bundle being completely separated and the helices unraveling. However, the complex is still bound, and the binding contacts are only partly affected by the unfolding monomer. In this case, the change in COM distance is due mainly to monomer unfolding, and not complex dissociation.

4.3 Discussion

The results show that two main pathways compete: monomer unfolding and complex dissociation. For a homodimer, as used in this particular study, a symmetric charge partitioning produces the largest intermolecular repulsion and the minimum total potential energy. The intermolecular repulsion works to lower the barrier for dissociation. Thus, for charge partitionings near symmetric, one expects that dissociation will be favored more than monomer unfolding. Alternatively, for very asymmetric charge partitioning, one monomer carries the bulk of the charge. In this limit, the intermolecular repulsion is near a minimum while the intramolecular repulsion is strong and in turn works to lower the barrier for monomer unfolding. Thus, for charge partitionings very far from symmetric, one expects that monomer unfolding will be favored more than complex dissociation. These trends are completely consistent with the Coulomb repulsion model (Section 3.3.1) and are supported by the present calculations.

From this point of view, the behavior of the system should be properly viewed on a two dimensional free energy surface with one dimension representing a coordinate for complex dissociation, such as the distance between the residues in the two monomers involved with binding the dimer, and the other representing a coordinate for monomer unfolding, such as the mean number of overcrossings. Figure 4.6 represents schematic views of such a free energy surface in these two different limiting cases. Both figures were constructed by considering a sum of two different independent energy terms. One term was an increasing function in the monomer unfolding coordinate. The other term was a gaussian function in the complex dissociation coordinate. For simplicity, each of these terms was approximated as being independent of the other, that is monomer unfolding does not significantly affect the binding of the complex, and vice versa. The only difference in the construction of the plots of Figure 4.6 is the relative magnitudes of the monomer unfolding and complex dissociation terms.

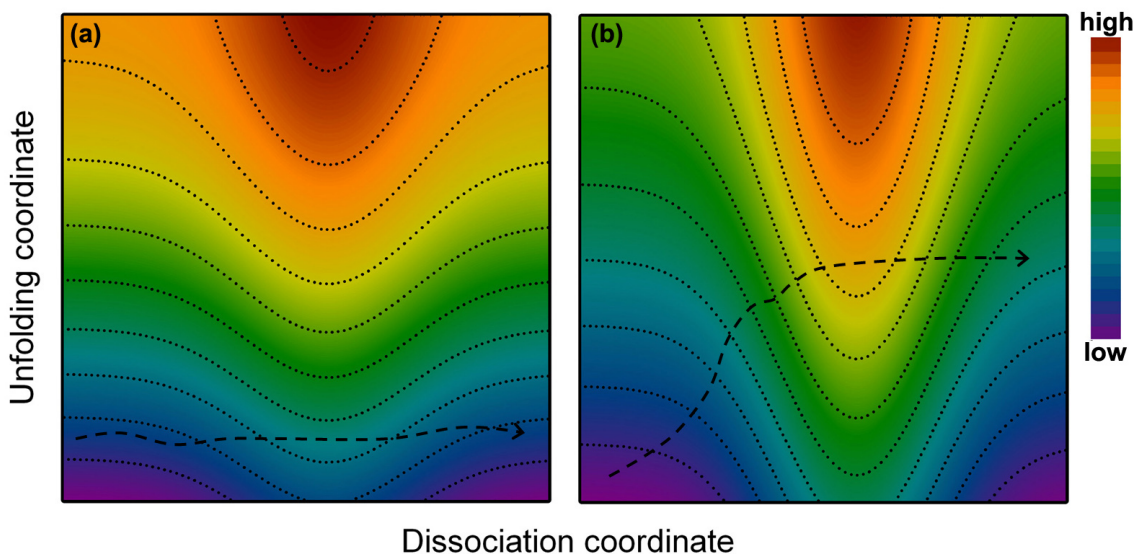


Figure 4.6: Schematic two dimensional free energy landscape as a function of complex dissociation and monomer unfolding coordinates for (left panel) a symmetric charge partitioning and (right panel) an asymmetric charge partitioning. The dashed lines represent possible paths for traversing from the bound dimer state, at the origin, to larger center of mass distances.

In Figure 4.6a, the complex dissociation barrier is lower than the monomer unfolding energy. This could represent the scenario in which the complex has a charge partitioning that is close to symmetric. In the constraint MD calculations, the centers of mass of the monomers are forced to separate. Examining Figure 4.6a shows that the lowest energy path is to move almost directly along the complex dissociation coordinate over the barrier, as indicated by the dashed line. In this case, one would predict that the complex dissociates promptly into two fragments with little unfolding during the process. In other words, the change in the complex dissociation coordinate is synonymous with the change in COM distance, and the latter represents a reasonable reaction coordinate for dissociation. This is precisely the behavior seen for the symmetric charge partitioning systems in the present results.

In Figure 4.6b, the monomer unfolding energy is lower than the complex dissociation barrier. This could represent the scenario in which the complex has a charge partitioning that is quite asymmetric. In this particular case, the contours guide the system along the monomer unfolding coordinate. However, the MD calculations force the COM distance to increase. As the COM distance is increased, the contours of the landscape force the system to move further along the monomer unfolding coordinate. That is, the simulations force the monomer to unfold, all the while climbing a higher and higher energy terrain, as

indicated by the path of the dashed line. This manifests itself in Figure 4.2 as a greatly increasing free energy barrier. In essence, the molecular dynamics algorithm is pulling on the dimer but the higher charged monomer would rather unfold than break the contacts binding the complex together. The increase in free energy seen in Figure 4.2 is thus the result of moving more along the monomer unfolding coordinate rather than the dissociation one, and the COM distance has more of the character of this unfolding coordinate than a complex dissociation one. However, because Figure 4.2 is a function only of the COM distance, the full two-dimensional nature of the free energy surface is not directly apparent. It must be inferred.

It should be emphasized that the balance between inter- and intra-molecular Coulomb energies determines the dissociation path of the protein complex. In other words, charge partitioning plays a major role in the protein complex dissociation mechanism.

The arguments developed above for homodimeric complexes can be extended to multimeric complexes as well. Consider a multimeric complex with n monomers, and for simplicity imagine all the monomers to be the same, and that the complex formed from a native solution with total charge Q . In practice, such complexes are formed with large total charges but small charges per monomer, Q/n . This implies that the intramolecular repulsion within each monomer is small, hence the unfolding barrier should be large. Additionally, if the complex dissociates by losing a single monomer, that monomer begins with a charge of Q/n while the remaining part of the complex has a charge of $Q(n-1)/n$. In other words, the charge is partitioned in a very asymmetric manner among the dissociating fragments, so that the intermolecular repulsion between these fragments should be small. The dissociation barrier for ejecting such a monomer should also be large. Overall then, the barriers for both monomer unfolding and monomer ejection should be high when the multimeric complex is in its ground state. The schematic free energy surface might be represented by that in the Figure 4.6b.

Now consider the effect of introducing energy into the complex, such as by collisional activation. Because in the ground state there is a distribution of charge partitioning expected in the complex (Section 3.2.2), it is probable that one monomer will have a charge slightly higher than average. For this monomer, the barriers represented in Figure 4.6 will be lower than are the ones for the other monomers, and thus will be the first to be affected. Recent surface-induced dissociation (SID) experiments of Wysocki and coworkers [35] provide supporting evidence for this picture. In this work, SID induces protein complex dissociation quickly compared with the timescales for other motions. It was found that protein complexes eject monomers with charges consistent with a uniform distribution. For example, a tetramer ejects a monomer carrying away approximately 1/4 of the total charge. These experiments show that in the ground state, the charges in the complex are generally arranged in a uniform manner, with some fluctuations about this limit.

Returning now to the contours in Figure 4.6b, the system will move along the monomer unfolding direction as energy is added. Assuming that charges are mobile under these conditions, an unfolding monomer will sequester additional charges in order that the surface charge density maintains an approximately constant value. However, as the charge on this monomer increases, the intramolecular repulsion increases AND the charge partitioning between the monomer and the rest of the complex becomes more symmetric thereby increasing the intermolecular repulsion. That is, the flow of charge on to a partially unfolding monomer causes a lowering of both the monomer unfolding and monomer ejection barriers.

In essence, the free energy terrain continues to look like that in Figure 4.6b but is generally lowered everywhere. This lowering then opens up additional energy-allowed pathways and the system continues to move along both the unfolding and dissociation pathways. In doing so, additional unfolding of the monomer results, which again leads to additional charges being sequestered, which again leads to an overall lowering of the unfolding and dissociation barriers. This process continues until eventually the system passes over the dissociation barrier and the monomer is ejected. Notice that in principle this process should continue until the monomer and remaining complex attain a symmetric charge partitioning, each having a charge of $Q/2$. However, in order to reach this limit, the monomer must unfold to the point where its surface area is about the same as that of the remaining complex. This may not be possible in all cases.

As detailed previously (Section 3.3), this argument is consistent with the reported experimental results involving the dissociation of multimeric protein complexes. It is also consistent with the most recent experiments of Wysocki and coworkers [35] who find that when using collision-induced dissociation, a variety of protein complexes decay by ejecting a monomer that carries away close to 50% of the charge.

The arguments above indicate that the large total charges in multimeric protein complexes greatly favor the unfolding and ejection of a monomer. However, these arguments do not produce the same result when applied to homodimeric complexes, since any motion of charge preferentially on to one monomer should cause the dissociation barrier to increase. However, several research groups [19, 33] have observed predominant asymmetric dissociation pathways during the dissociation of homodimers with low charges under some experimental conditions. The results of Jones *et al.* [9] are particularly striking since they show that cytochrome c dimer produced under the same conditions dissociates into a predominantly symmetric channel with SID but an asymmetric one with CID.

If the SID spectrum represents the ground state distribution of charge in the complex, then it shows that a symmetric charging is present, consistent with the Coulomb repulsion model. To rationalize the CID results (a lower energy technique than SID) requires that charges migrate to produce an asymmetric partitioning, presumably with a commensurate unfolding or partial unfolding of the higher charged monomer. However, according

to the calculations reported herein, this asymmetric state should have a higher barrier to dissociation than the symmetric one. How is this to be resolved?

First, to date experiments that have reported asymmetric dissociation for homodimers have involved complexes with small total charges. In this limit, the Coulomb repulsion model could fail because the net charges do not dominate the interactions enough. Second, the present calculations use static charge configurations for all MD trajectories. Charge motion is not explicitly included. Instead, the effect of charge transfer can be inferred by imagining that the system hops from one charge partitioning surface to another. It is possible that pathways which couple charge transfer and unfolding may exist that are not revealed by fixed charge configuration trajectories. Such pathways might involve lower free energies. Third, in the current calculations and with the force field employed, the number of charges on the monomer was always smaller than that needed to completely overcome the unfolding barrier. That is, if these charges were placed on a single monomer, it would remain mostly folded. Thus, even for the asymmetric partitionings, it requires energy to unfold a monomer and this accounts in part for the ever increasing free energy barriers seen in Figure 4.2 for these cases. For other protein homodimers, it may be the case that the barrier to unfolding is lower than the dissociation barrier, or the unfolding pathway is composed of a succession of intermediate states, each of which is accessed by low barriers. In this case, the system moves first along the unfolding coordinate, gradually gaining energy until complex dissociation finally occurs. Such an incremental process may be more favorable than a direct dissociation pathway from the symmetric charge partitioned state, especially if entropy gained from the former unfolding pathway lowers the free energy.

Chapter 5

Dissociation Coordinates

Properties related to the dissociation of dimers were analyzed in order to obtain a better understanding of when and how dissociations of charged protein complexes occur. More specifically, two dissociation parameters related to the number of intermolecular hydrogen bonds and intermolecular residual distances were defined and calculated. These parameters were then compared with free energy barriers. In addition, components of the potential energy function were also calculated as a function of COM distance.

5.1 Method

The D10 charge state (with a total charge of +10 on the dimer) with charge partitionings of M5/M5, M7/M3, M8/M2 was selected for this study. For each charge partitioning, 10 configurations selected for free energy calculations in Chapter 4 were chosen for further analysis. For each charge configuration, five trajectories with different initial velocities were analyzed. That is for this study we used the same output trajectories produced in free energy calculations as a function of COM distance.

5.2 Results

In order to evaluate the dissociation process, the number of intermolecular hydrogen bonds between the monomers was calculated as a function of COM distance. The number of intermolecular hydrogen bonds were calculated using a utility program in the Gromacs package. In this program, all intermolecular donor-accepter pairs are tested and a hydrogen bond is accepted if the hydrogen-accepter distance is within 3.5 Å and acceptor-donor-hydrogen angle is within 30°. The number of intermolecular hydrogen bonds were counted at 0.1 ps time intervals during the constraint MD run and then averaged over the 100 ps time period. For a representative charge configuration of the M5/M5 and M7/M3 charge partitionings, the average number of intermolecular hydrogen bonds as a function of COM distance for 5 trajectories are shown in Figure 5.1.

Figure 5.1(a) shows that the number of intermolecular hydrogen bonds remains about constant on average until all the bonds have been dissociated near the 6 nm COM distance for all 5 trajectories. However, this was not the case for all charge configurations. For some

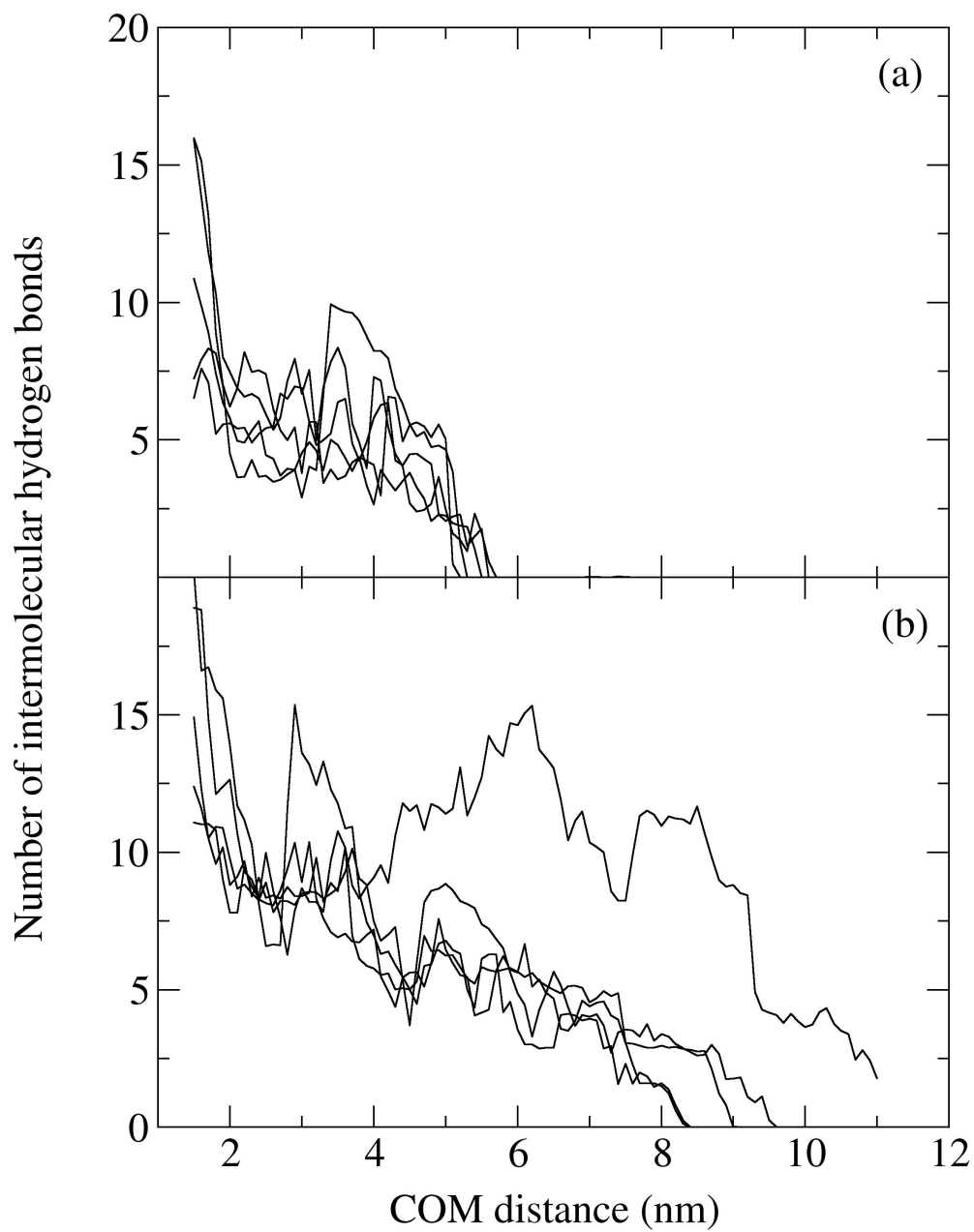


Figure 5.1: The average number of intermolecular hydrogen bonds as a function of COM distance for five different trajectories started from the same charge configuration in the (a) M5/M5 and (b) M7/M3 charge partitionings.

trajectories, a few intermolecular hydrogen bonds remained intact within the simulated COM distance range. For example, Figure 5.1(b) shows that a few intermolecular hydrogen bonds still remain at the highest simulated COM distance for one of the 5 trajectories. Even though few trajectories remained intact with same inter molecular hydrogen bonds, all hydrogen bonds were broken in the majority of trajectories (more than 90% of the trajectories for each charge partitioning) within the simulated COM distance range. Unlike the behavior for the symmetric charge partitioning in Figure 5.1(a), the asymmetric ones in Figure 5.1(b) show the number of hydrogen bonds gradually breaking in a linear fashion with COM distance until the final dissociation point.

Let R_{hb} be the COM distance at which the last intermolecular hydrogen bond dissociates. Here, we consider the breaking of the last hydrogen bond when the average number of intermolecular hydrogen bonds becomes less than 0.5. The R_{hb} values could be determined only for the trajectories in which all intermolecular hydrogen bonds have broken within the simulated COM distance. The R_{hb} values were averaged over all trajectories for each charge partitioning (i.e. M5/M5, M7/M3 and M8/M2). Averaged R_{hb} values along with their standard deviations for the M5/M5, M7/M3 and M8/M2 charge partitionings as well as for the neutral dimer are listed in Table 5.1. In addition to these values, the COM distances corresponding to the peak of the free energy barrier taken from Figure 4.2 in Chapter 4 are given in Table 5.1.

Charge Partitioning	R_{hb} distance (nm)	peak of free energy barrier (nm)
Neutral	6.33 ± 0.75	6.7
M5/M5	5.69 ± 0.53	5.3
M7/M3	7.22 ± 1.30	6.10
M8/M2	11.90 ± 3.63	13.30

Table 5.1: For differing charge partitionings, this table compares the average distance that the last intermolecular hydrogen bond between two monomers breaks, R_{hb} , with the COM position of the peak of the corresponding calculated free energy barrier from Figure 4.2.

For all charge partitionings, there is a good correlation between the COM distances at the peak of free energy barriers and the average R_{hb} distances. Furthermore, the breaking of intermolecular hydrogen bonds gives a direct measure of the dissociation process since we always expect hydrogen bonds to exist between monomers in the bound complex. Therefore, R_{hb} tracks the complete breaking of contact between the monomers and defines well the point of complex dissociation. The fact that it directly corresponds to the peak of the free

energy barrier suggests that the COM distance is a reasonable coordinate for describing the dissociation process.

As shown in Table 5.1, the R_{hb} distance increases as the charge partitioning moves from symmetrical to asymmetrical. These results imply that asymmetrically charged dimers exist for longer COM distances than symmetrically charged ones. For the M8/M2 charge partitioning, even at a higher COM distance, the two monomers stick together by one or more non-covalent bonds because the higher charged monomer begins to unfold before dissociation (Chapter 4). Preferential charging of a monomer leads to a higher intramolecular Coulomb repulsion and lowers the unfolding barrier. Therefore, the increase of the COM distance in this case leads towards unfolding instead of dissociation.

However, the symmetrically charged dimers (M5/M5) dissociate at lower COM distances without unfolding (Chapter 4). The same trend can also be seen for the neutral dimer. The R_{hb} distance of the neutral dimer is larger than that of the M5/M5 charge partitioning, because intermolecular Coulomb repulsion lowers the dissociation barrier for the latter. However, the R_{hb} values of asymmetric charge partitionings are higher than that of the neutral dimer because unfolding of higher charged monomer leads to a larger R_{hb} value.

Intermolecular residual distances can also be a good descriptor for the dissociation of protein complexes. Thus, an additional parameter is described below that considers the distances between residues, one in each monomer. The COM coordinates for all residues were first calculated and then the distance between the centers of mass of two residues, r , was calculated for all possible intermolecular residual pairs (one residue in each monomer) in the system at 0.1 ps intervals during the constraint MD run. This gave 17161 intermolecular residual distances at each interval. Our goal is to define a parameter, R_{rd} , that represents the minimum intermolecular residual distance over the entire trajectory. In other words, at each time, we want R_{rd} to be the smallest value such that (on average) one intermolecular residual pair distance is within R_{rd} . In our case, each trajectory was sampled 1000 times for residual pair distance calculations and all the r distances from the entire trajectory were collected and ordered from smallest to highest. The lowest 1000 values were selected and among them, the largest one was set to R_{rd} . This procedure guarantees that on average at each time, the probability of having at least one intermolecular residual distance less than R_{rd} is unity.

The R_{rd} values were averaged over all 50 trajectories at each COM distance. Distributions of average R_{rd} as a function of COM distance for the M5/M5, M7/M3, M8/M2 charge partitionings and the neutral dimer are shown in Figure 5.2. The bars in Figure 5.2 show the standard deviation of the average R_{rd} value calculated from the 50 trajectories, and the dashed vertical lines indicate the corresponding R_{hb} values from Table 1. Furthermore, red dashed lines with slope one represent the change in R_{rd} value as a function of COM distance expected if monomers dissociate without undergoing further structural changes. In other

words, along this line the increase in R_{rd} is the same as the increase in COM distance after dissociation of the monomers. If the increasing rate of the average R_{rd} value is less than that of the COM distance (i.e. slope < 1), we expect that one or both monomers unfold or stretch after breaking all contacts between the monomers. On the other hand, if the increasing rate of the R_{rd} value is higher than that of the COM distance (i.e. slope > 1), one or both monomers are compressed.

Consulting Figure 5.2 shows that for all charge partitionings and the neutral one, the average R_{rd} values stay almost constant up to a certain COM distance and then increase sharply. The flat areas imply that at least one intermolecular residual pair stays non-covalently bonded throughout this COM distance range. On the other hand, a sharp increase of the R_{rd} value indicates the breaking of all non-covalent bonds and complete dissociation. Due to the large standard deviations in Figure 5.2, it is hard to define the distance at which dissociation occurs, hence the reason for calculating R_{hb} values. However, as described above, the R_{rd} values provide a better understanding of structural changes occurring after dissociation.

It would be worth exploring whether any specific attached intermolecular residual pair or pairs exist at the COM distance when dimer dissociation occurs. Thus, we analyzed the trajectories for the M8/M2 charge partitioning at the COM distances where the R_{rd} value starts to increase. At this COM distance, we screened intermolecular residual pairs with $r \leq 1$ nm and essentially determined the residue pair corresponding to the last broken intermolecular hydrogen bond. The analysis shows that the N-terminus of the monomer with the higher charge was always included in the selected intermolecular residual pairs, indicating that the N-terminus is preferentially attached to the other monomer in the M8/M2 charge partitioning. However, the residue paired with the N-terminus is not always the same. Furthermore, the N-terminus of the higher charged monomer was preferentially protonated for all charge configurations in the M8/M2 charge partitioning. Thus, this protonated N-terminus always bonds to the other monomer. However, the N-terminus is not preferentially charged when monomers have fewer charges. For example, in the M5/M5 charge partitioning, each monomer has fewer charges and there are other favorable charge configurations excluding the N-terminus. The same analysis was repeated for the M5/M5 charge partitioning and indicated that no specific residue or residual pair is preferentially attached at the COM distance when the dimer dissociates.

It is important to investigate how each term in the OPLS potential energy function (angle bending potential, dihedral angle potential, intra- and intermolecular Lennard-Jones potentials, and intra- and intermolecular Coulomb potentials) changes as a function of COM distance. It should be noted that the bond stretching contribution to the total potential energy is zero because all bonds are constrained in all MD runs. The energy of a selected term was calculated at 0.1 ps time intervals and averaged over a 100 ps time period at each

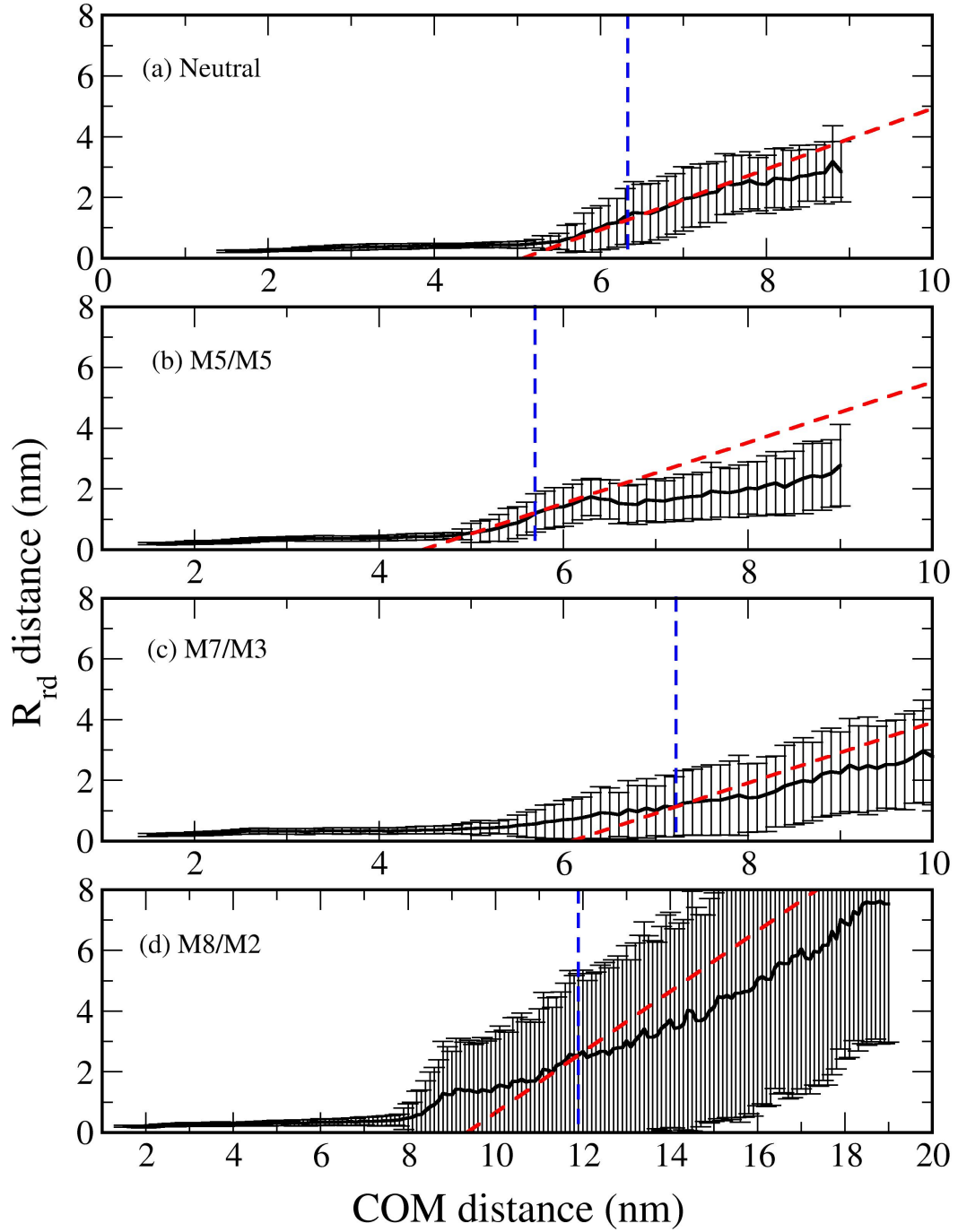


Figure 5.2: Average R_{rd} distances as a function of COM distance for the (a) Neutral, (b) M5/M5, (c) M7/M3 and (d) M8/M2 charge partitionings (solid line). Bars indicate the standard deviation of the average. The dashed blue vertical lines are at the corresponding R_{hb} values from Table 5.1. The red dashed lines represent the idealized change of R_{rd} when two monomers separate without structural changes (slope = 1).

fixed COM distance. For each trajectory, the resulting energies were then shifted by setting the energy of this term to zero at the COM distance of the bound dimer. In other words, only energy differences relative to the ground state dimer were calculated for each potential term. Relative energies for all trajectories in the same charge partitioning were averaged at each charge configuration. The above steps were repeated with all potential energy terms in every charge partitioning and the relative values of each term were plotted as a function of COM distance in Figure 5.3.

5.3 Discussion

Together, the various calculated properties present a consistent picture for the sequence of events that occur during dimer dissociation. Consider first the neutral cluster. The intermolecular hydrogen bond data shows that over a fairly range of values, the dimer dissociates when the COM distance is about 6.3 nm. However, Figure 5.2(a) shows that the R_{rd} values stay approximately constant and at small values up to this COM distance. This means that as the COM of the monomers is increased, the binding region remains virtually unchanged. In other words, the monomers are being stretched, much like an elastic, in this region. This stretching process is both energetically unfavorable (as seen by the increase in the total potential energy within this COM range in Figure 5.3(a)) and entropically unfavourable due to a lowering of the conformational flexibility in the monomers. As a result, the free energy increases.

As the COM is increased further, the hydrogen bonds within the binding region are no longer able to oppose the stretching forces in the monomers, and the dimers dissociate at a well defined COM distance, shown by the vertical dashed line in Figure 5.2(a). Note that the values of R_{rd} in Figure 5.2(a) start to increase at a COM distance a bit smaller than the R_{hb} value. This happens because the dissociation occurs over a small range of COM distances, so some trajectories have already dissociated before R_{hb} is reached, and some dissociate after this value. According to Table 1, this range is about 0.75 nm on either side of R_{hb} . Since the R_{rd} values are averages over *all* trajectories, the early dissociating ones start to increase the value of R_{rd} at COM distances before R_{hb} . At about the value of R_{hb} , the R_{rd} values increase linearly with unit slope in Figure 5.2(a), indicating that the monomers are not changing in size after dissociating. Note that this also implies the stretching occurring in the monomers before dissociation is not undone after dissociation.

In fact, examining the potential energy values in Figure 5.3(a) shows that indeed the monomers start to relax immediately after dissociation. This relaxation does not produce an overall change in size of the monomers instantly. However, it will be noticed at a COM distance of approximately 8.0 nm in Figure 5.2(a), that the average R_{rd} values fall below the curve with unit slope, implying that the sizes of the monomers are increasing. Overall, the

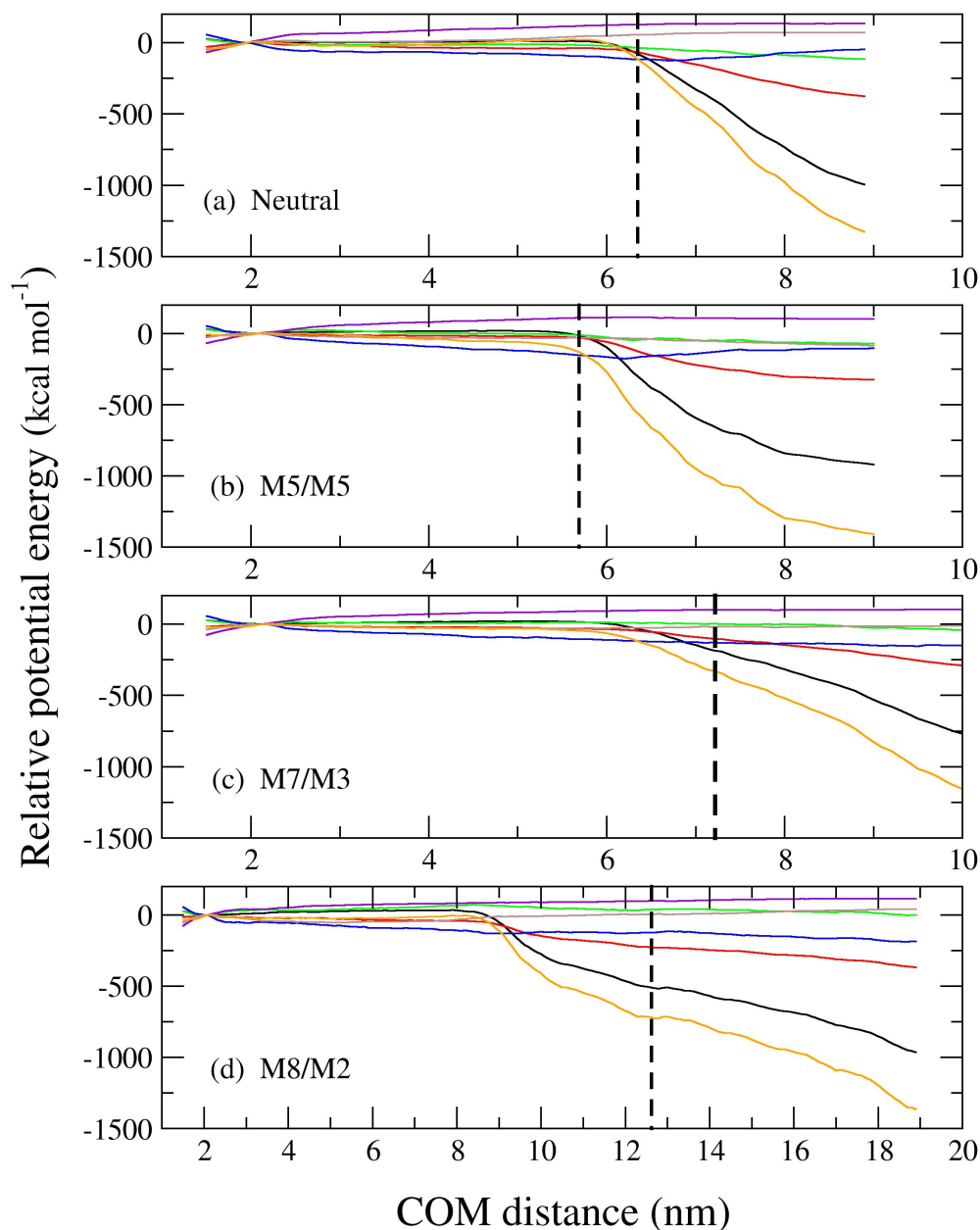


Figure 5.3: Changes in the potential energy components (angle bending (black), dihedral angle (red), intramolecular Lennard-Jones (green), intramolecular Coulomb (blue), intermolecular Lennard-Jones (violet), intermolecular Coulomb (brown)) and total potential energy (orange) as a function of COM distance for the (a) neutral, (b) M5/M5, (c) M7/M3 and (d) M8/M2 charge partitionings. Vertical dashed lines indicate the corresponding values of R_{hb} from Table 5.1.

relaxation that sets in immediately after dissociation takes some time before manifesting itself in a more loosely structured monomer. Our simulations at each COM distance (of 100 ps duration) are too short to fully account for this relaxation, and in fact, it continues to progress as the COM distance is increased (since the initial conditions for the next COM distance simulation are taken from the final values at the previous distance). From the trends seen in Figure 5.2, it would require simulation times approximately an order of magnitude longer if one wanted to allow the full relaxation to occur at each COM distance.

Overall then, the neutral dimer dissociates as essentially two folded monomers. These monomers relax after dissociation and expand somewhat but generally retain their bound state forms.

The behavior of the M5/M5 charge partitioning exhibits many of the same characteristics as the neutral dimer except for two differences. First, the R_{hb} value is smaller because the monomers are now charged. These charges lead to greater intermolecular repulsion which lowers the barrier to dissociation and moves it to smaller COM distances. Second, the deviation of the increase of R_{rd} away from the line of unit slope starts at a COM distance of approximately 6.5 nm in Figure 5.2(b), a smaller value than for the neutral dimer. This indicates that the monomers relax faster, and this is not unexpected. The monomers have +5 charges each, and this increases the intramolecular repulsion compared with a neutral monomer. This increased repulsion facilitates the relaxation and swelling of the monomer after dissociation.

The behavior of the M7/M3 charge partitioning starts to exhibit behaviour that differs from that of the neutral or lower charge partitionings. Consider first the increase of the R_{rd} values in Figure 5.2(c). Unlike the neutral and M5/M5 charge partitioning which saw these values increase with unit slope, the values in the M7/M3 case increase with a slope less than unity. This implies that a monomer is increasing in size as dissociation occurs. In this case, there are +7 charges on one monomer, and as the COM distance increases, the repulsion between these charges begin to facilitate monomer expansion. The result is that the value of R_{hb} exceeds that of the neutral dimer somewhat, indicating that at the point of dissociation, the higher charged monomer has increased in radius by about 1 nm compared with the neutral one. If you like, the relaxation that occurs after dissociation in the M5/M5 case starts to begin at or slightly before dissociation in the M7/M3 case. This is evidenced by the potential energy curves in Figure 5.3(c) which show the drop in energy occurring slightly before R_{hb} . The dissociation process itself is also less prompt. As seen in Figure 5.1(b), the intermolecular hydrogen bonds are not all broken simultaneously, as in the neutral and M5/M5 cases. Instead, there is an approximately linear decay in the number of such bonds, indicating that the departing monomer is slowly peeled away, rather than breaking promptly.

This behavior is even more pronounced for the M8/M2 partitioning. In this case, one

monomer has +8 charges, and is quite susceptible to structural changes. As the COM distance is increased, this monomer begins to expand well before dissociation, as evidenced by the R_{rd} values in Figure 5.2(d). The R_{hb} values are subsequently pushed to large values because the unfolding of the monomer delays the onset of dissociation. Not until the monomer is sufficiently extended will enough force be exerted on the binding region to cause dissociation.

There is another factor also at work for the higher charged states. With higher charges comes a higher probability that the N-terminus is protonated, and in the present case, all charge configurations for the M8/M2 partitioning contained a protonated N-terminus. Additionally, this N-terminus is always the last contact to break in the dissociation process. This helps facilitate monomer unfolding by tethering one end of the monomer chain firmly in place. Thus, the forces exerted by the ever increasing COM distance naturally work to unfold the monomer. In this sense, the situation is similar to single molecule pulling experiments that unfold a protein by exerting mechanical forces at one end of the molecule while the other end is attached to a surface [109].

Chapter 6

Concluding Remarks

The main focus of this thesis work was to obtain a better understanding of charge separation during the dissociation of protein complex ions in the gas phase. This work was conducted with three studies using the cytochrome c' dimer as the model system. These studies included potential energy calculations of bound dimers using a simple electrostatic model and molecular dynamics simulations, free energy calculations of dissociation using COM constraint molecular dynamics simulations, and analysis of the change in intermolecular properties during the dissociation process. Findings from these studies led to more detailed understanding of the charge distributions among fragments as well as the dissociation mechanism of protein complexes, as detailed in the discussion sections in each chapter.

Results from this thesis work showed that the dissociation process of charged multimeric protein complexes can be explained by an electrostatic model based upon the repulsion between the net charges on the protein. This model, when extended to protein complexes, predicts that charges should seek to arrange themselves so as to maintain approximately a uniform surface charge density. Furthermore, no single charge configuration will be important but rather distributions of configurations will contribute to the behavior of the system. When considering the dissociation mechanism, several dissociation paths are available for charged protein complexes. These paths depend upon two competitive processes: dissociation and monomer unfolding. The lowest energy barrier for dissociation always occurs when both fragment ions have the same charge, that is the total charge is divided symmetrically among the fragment ions. Alternatively, an asymmetric partitioning of charges among the dissociating fragments preferentially lowers the unfolding barrier relative to the dissociation one.

For multimeric complexes, it is favorable to have a single monomer gradually unfold and sequester charges. This process works cooperatively to decrease both the barriers for monomer unfolding and monomer ejection. On the other hand, the Coulomb repulsion model will be poorer for complexes with low charges. For an example, the Coulomb repulsion model is unable to explain the asymmetric charge distribution during the dissociation of protein dimers reported under some conditions. In order to account for some particular experimental observations of the predominance of asymmetric dissociation channels, clearly, more detailed investigations are needed.

In this thesis work, all calculations were conducted with static charges. Specifically

in the free energy calculations, charge distributions were selected to be the lowest energy charge distributions in the bound dimer state. However the selected distributions may not be the lowest energy charge distributions for other distances. In other words, the lowest energy charge configurations may change with COM distance. Furthermore, a number of experimental measurements [12, 18, 102–105] as well as some theoretical calculations [106–108] indicate that charge transfer in proteins is labile. Therefore, it is necessary to include the motion of charges in these calculations if we want to account for these effects and explore other dissociation channels. For this purpose, a special algorithm should be developed to move protons (hydrogen atoms) between protonation sites after a certain number of MD steps in order to have the lowest energy charge configuration for relaxed structures. This process should be repeated until the energy converges to a lower value.

On the other hand, as discussed in Chapter 4, the free energy landscape of dissociation vs. unfolding can be used to explore the available dissociation paths. Therefore, it is important to create a free energy landscape for each charge partitioning with different degrees of unfolding of the monomers by the same procedure used in Chapter 4. In other words, several free energy profiles of dissociation for a selected charge partitioning are obtained with different degrees of unfolding and then used to create a contour map which is similar to Figure 4.6. However, the computational effort involved in doing this was too great. A possible way to do this calculation is to perform coarse-grained molecular dynamic simulations in which a system is represented by a reduced number of degrees of freedom in comparison with an all-atom description in molecular dynamic simulations, and the computational effort is reduced considerably.

One of the most important problems facing free energy calculations by computer simulations for complex systems such as proteins is the need to enhance the search of their configurational space. Especially in higher temperature molecular dynamics simulations, the additional kinetic energy available enhances the probability of visiting different barrier regions or helps the system cross higher energy barriers. Thus, the system can access regions that are not accessible with short finite time simulations at room temperature. It would also be interesting to examine the temperature dependence of the free energy barrier. Examining the temperature dependence could be used to extract kinetic information such as Arrhenius activation parameters. It should be noted that a better sampling technique is required if one wants to get kinetic quantities from MD simulations.

Overall, this thesis work demonstrates that the dynamics of protein complex ions, such as structural relaxation and unfolding, play a significant role in the dissociation mechanism. These dynamic details are difficult to obtain by experimental studies and furthermore, not accessible by static structure calculations. Molecular dynamic simulation has therefore gained much attention as a valuable tool for performing these types of studies.

References

- [1] Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R., Principles and practice of electrospray ionization-mass spectrometry for large polypeptides and proteins, *Mass Spectrom Rev*, **1991**, *10*, 359–452.
- [2] Cole, R. B., *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications*; Wiley: New York, **1997**.
- [3] Veenstra, T. D., Electrospray ionization mass spectrometry in the study of biomolecular non-covalent interactions, *Biophys Chem*, **1999**, *79*, 63–79.
- [4] Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, *J Mass Spectrom*, **2000**, *35*, 804–817.
- [5] Loo, J. A., Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes, *Int J Mass Spectrom*, **2000**, *200*, 175–186.
- [6] Schwartz, B. L.; Bruce, J. E.; Anderson, G. A.; Hofstadler, S. A.; Rockwood, A. L.; Smith, R. D.; Chilkoti, A.; Stayton, P. S., Dissociation of tetrameric ions of non-covalent streptavidin complexes formed by electrospray ionization, *J Am Soc Mass Spectrom*, **1995**, *6*, 459–465.
- [7] Heck, A. J. R.; van den Heuvel, R. H. H., Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom Rev*, **2004**, *23*, 368–389.
- [8] Benesch, J. L. P.; Robinson, C. V., Mass spectrometry of macromolecular assemblies: preservation and dissociation, *Curr Opin Struct Biol*, **2006**, *16*, 245–251.
- [9] Jones, C. M.; Beardsley, R. L.; Galhena, A. S.; Dagan, S.; Cheng, G.; Wysocki, V. H., Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions, *J Am Chem Soc*, **2006**, *128*, 15044–15045.
- [10] Light-Wahl, K. J.; Schwartz, B. L.; Smith, R. D., Observation of the noncovalent quaternary associations of proteins by electrospray ionization mass spectrometry, *J Am Chem Soc*, **1994**, *116*, 5271–5278.

- [11] Versluis, C.; Heck, A. J. R., Gas-phase dissociation of hemoglobin, *Int J Mass Spectrom*, **2001**, *210-211*, 637–649.
- [12] Felitsyn, N.; Kitova, E. N.; Klassen, J. S., Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior, *Anal Chem*, **2001**, *73*, 4647–4661.
- [13] Sobott, F.; McCammon, M. G.; Robinson, C. V., Gas-phase dissociation pathways of a tetrameric protein complex, *Int J Mass Spectrom*, **2003**, *230*, 193–200.
- [14] Benesch, J. L. P.; Sobott, F.; Robinson, C. V., Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry, *Anal Chem*, **2003**, *75*, 2208–2214.
- [15] van den Heuvel, R. H.; Heck, A. J., Native protein mass spectrometry: from intact oligomers to functional machineries, *Curr Opin Chem Biol*, **2004**, *8*, 519 – 526.
- [16] Loo, J. A.; Berhane, B.; Kaddis, C. S.; Wooding, K. M.; Xie, Y.; Kaufman, S. L.; Chernushevich, I. V., Electrospray ionization mass spectrometry and ion mobility analysis of the 20s proteasome complex, *J Am Soc Mass Spectrom*, **2005**, *16*, 998 – 1008.
- [17] Benesch, J. L. P.; Aquilina, J. A.; Ruotolo, B. T.; Sobott, F.; Robinson, C. V., Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies, *Chem Biol*, **2006**, *13*, 597–605.
- [18] Sinelnikov, I.; Kitova, E. N.; Klassen, J. S., Influence of coulombic repulsion on the dissociation pathways and energetics of multiprotein complexes in the gas phase, *J Am Soc Mass Spectrom*, **2007**, *18*, 617–631.
- [19] Jurchen, J. C.; Williams, E. R., Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers, *J Am Chem Soc*, **2003**, *125*, 2817–2826.
- [20] Csiszar, S.; Thachuk, M., Using ellipsoids to model charge distributions in gas phase protein complex ion dissociation, *Can J Chem*, **2004**, *82*, 1736–1744.
- [21] Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B., Molecular beams of macroions, *J Chem Phys*, **1968**, *49*, 2240–2249.
- [22] Yamashita, M.; Fenn, J. B., Electrospray ion source. Another variation on the free-jet theme, *J Phys Chem*, **1984**, *88*, 4451–4459.

- [23] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules, *Science*, **1989**, *246*, 64–71.
- [24] Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V., Protein complexes in the gas phase: technology for structural genomics and proteomics, *Chem Rev*, **2007**, *107*, 3544–3567.
- [25] Iribarne, J. V.; Thomson, B. A., On the evaporation of small ions from charged droplets, *J Chem Phys*, **1976**, *64*, 2287–2294.
- [26] Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal Chem*, **1988**, *60*, 2299–2301.
- [27] Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and polymer analyses up to m/z 100000 by laser ionization time-of-flight mass spectrometry, *Rapid Commun Mass Spectrom*, **1988**, *2*, 151–153.
- [28] Griffiths, W. J.; Jonsson, A. P.; Liu, S.; Rai, D. K.; Wang, Y., Electrospray and tandem mass spectrometry in biochemistry, *Biochem J*, **2001**, *355*, 545–561.
- [29] Dawson, P. H., *Quadrupole mass spectrometry and its applications*; Elsevier, Amsterdam, **1976**.
- [30] Wollnik, H., Time-of-flight mass analyzers, *Mass Spectrom Rev*, **1993**, *12*, 89–114.
- [31] March, R. E., An introduction to quadrupole ion trap mass spectrometry, *J Mass Spectrom*, **1997**, *32*, 351–369.
- [32] Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer, *Mass Spectrom Rev*, **1998**, *17*, 1–35.
- [33] Versluis, C.; van der Staaij, A.; Stokvis, E.; Heck, A. J. R.; de Craene, B., Metastable ion formation and disparate charge separation in the gas-phase dissection of protein assemblies studied by orthogonal time-of-flight mass spectrometry, *J Am Soc Mass Spectrom*, **2001**, *12*, 329–336.
- [34] Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry, *J Mass Spectrom*, **2004**, *39*, 1091–1112.
- [35] Wysocki, V. H.; Jones, C. M.; Galhena, A. S.; Blackwell, A. E., Surface-induced dissociation shows potential to be more informative than collision-induced dissociation for structural studies of large systems, *J Am Soc Mass Spectrom*, **2008**, *19*, 903–913.

- [36] Mohammed, S.; Chalmers, M. J.; Gielbert, J.; Ferro, M.; Gora, L.; Smith, D. C.; Gaskell, S. J., A novel tandem quadrupole mass spectrometer allowing gaseous collisional activation and surface induced dissociation, *J Mass Spectrom*, **2001**, *36*, 1260–1268.
- [37] Price, W. D.; Schnier, P. D.; Williams, E. R., Tandem mass spectrometry of large biomolecule ions by blackbody infrared radiative dissociation, *Anal Chem*, **1996**, *68*, 859–866.
- [38] Dunbar, R. C., BIRD (blackbody infrared radiative dissociation): evolution, principles, and applications, *Mass Spectrom Rev*, **2004**, *23*, 127–158.
- [39] Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B., Sustained off-resonance irradiation for collision-activated dissociation involving fourier transform mass spectrometry. Collision-activated dissociation technique that emulates infrared multiphoton dissociation, *Anal Chim Acta*, **1991**, *246*, 211–225.
- [40] Cooper, H. J.; Håkansson, K.; Marshall, A. G., The role of electron capture dissociation in biomolecular analysis, *Mass Spectrum Rev*, **2005**, *24*, 201–222.
- [41] Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process, *J Am Chem Soc*, **1998**, *120*, 3265–3266.
- [42] Chowdhury, S. K.; Katta, V.; Chait, B. T., Probing conformational changes in proteins by mass spectrometry, *J Am Chem Soc*, **1990**, *112*, 9012 – 9013.
- [43] van den Heuvel, R. H. H.; van Duijn, E.; Mazon, H.; Synowsky, S. A.; Lorenzen, K.; Versluis, C.; Brouns, S. J. J.; Langridge, D.; van der Oost, J.; Hoyes, J.; Heck, A. J. R., Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry, *Anal Chem*, **2006**, *78*, 7473–7483.
- [44] Geels, R. B. J.; van der Vies, S. M.; Heck, A. J. R.; Heeren, R. M. A., Electron capture dissociation as structural probe for noncovalent gas-phase protein assemblies, *Anal Chem*, **2006**, *78*, 7191–7196.
- [45] Felitsyn, N.; Kitova, E. N.; Klassen, J. S., Thermal dissociation of the protein homodimer ecotin in the gas phase, *J Am Soc Mass Spectrom*, **2002**, *13*, 1432–1442.
- [46] Ryce, S. A.; Wyman, R. R., Two sphere model for the asymmetric division of electrically charged liquid drops, *Can J Phys*, **1970**, *48*, 2571–2576.

- [47] Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs, *J Comput Chem*, **2005**, *26*, 1668–1688.
- [48] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J Comput Chem*, **1983**, *4*, 187–217.
- [49] MacKerel Jr., A. D.; Brooks, B.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M., CHARMM: the energy function and its parameterization with an overview of the program, In *The Encyclopedia of Computational Chemistry*, Vol. 1; John Wiley & Sons: Chichester, **1998**, 271–277.
- [50] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: a message-passing parallel molecular dynamics implementation, *Comput Phys Commun*, **1995**, *91*, 43–56.
- [51] Lindahl, E.; Hess, B.; van der Spoel, D., Gromacs 3.0: a package for molecular simulation and trajectory analysis, *J Mol Model*, **2001**, *7*, 306–317.
- [52] Gromacs user manual version 3.2. van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C.; www.gromacs.org, **2004**.
- [53] Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F., The GROMOS software for biomolecular simulation: GROMOS 05, *J Comput Chem*, **2005**, *26*, 1719–1751.
- [54] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable molecular dynamics with NAMD, *J Comput Chem*, **2005**, *26*, 1781–1802.
- [55] Mackerell Jr., A. D., Empirical force fields for biological macromolecules: overview and issues, *J Comput Chem*, **2004**, *25*, 1584–1604.
- [56] Alder, B. J.; Wainwright, T. E., Phase transition for a hard sphere system, *J Chem Phys*, **1957**, *27*, 1208–1209.
- [57] Norberg, J.; Nilsson, L., Advances in biomolecular simulations: methodology and recent applications, *Q Rev Biophys*, **2003**, *36*, 257–306.

- [58] McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins, *Nature*, **1977**, *267*, 585–590.
- [59] Cheatham, T. E.; Kollman, P. A., Molecular dynamics simulation of nucleic acids, *Annu Rev Phys Chem*, **2000**, *51*, 435–471.
- [60] Spieser, S. A. H.; van Kuik, J. A.; Kroon-Batenburg, L. M. J.; Kroon, J., Improved carbohydrate force field for Gromos: ring and hydroxymethyl group conformations and exo-anomeric effect, *Carbohydr Res*, **1999**, *322*, 264 – 273.
- [61] Tieleman, D. P.; Marrink, S. J.; Berendsen, H. J. C., A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems, *Biochim Biophys Acta, Rev Biomembr*, **1997**, *1331*, 235 – 270.
- [62] Ren, Z.; Meyer, T.; McRee, D. E., Atomic structure of a cytochrome c' with an unusual ligand-controlled dimer dissociation at 1.8 Å resolution, *J Mol Biol*, **1993**, *234*, 433 – 445.
- [63] Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., An all atom force field for simulations of proteins and nucleic acids, *J Comput Chem*, **1986**, *7*, 230–252.
- [64] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins: Struct , Funct , Bioinf*, **2006**, *65*, 712–725.
- [65] Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J Phys Chem B*, **2001**, *105*, 6474–6487.
- [66] Ponder, J. W.; Case, D. A., Force fields for protein simulations, In *Protein Simulations*; Daggett, V., Ed., Vol. 66 of *Advances in Protein Chemistry*; Academic Press, **2003**, 27 – 85.
- [67] MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J Phys Chem B*, **1998**, *102*, 3586–3616.

- [68] Jorgensen, W. L.; Tirado-Rives, J., The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *J Am Chem Soc*, **1988**, *110*, 1657–1666.
- [69] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A new force field for molecular mechanical simulation of nucleic acids and proteins, *J Am Chem Soc*, **1984**, *106*, 765–784.
- [70] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J Am Chem Soc*, **1996**, *118*, 11225–11236.
- [71] Hockney, R. W.; Eastwood, J. W., *Computer simulation using particles*; Taylor & Francis: Bristol, **1988**.
- [72] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath, *J Chem Phys*, **1984**, *81*, 3684–3690.
- [73] Nosé, S., A molecular dynamics method for simulations in the canonical ensemble, *Mol Phys*, **1984**, *52*, 255–268.
- [74] Nosé, S., A unified formulation of the constant temperature molecular dynamics methods, *J Chem Phys*, **1984**, *81*, 511–519.
- [75] Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions, *Phys Rev A*, **1985**, *31*, 1695–1697.
- [76] Woodcock, L. V., Isothermal molecular dynamics calculations for liquid salts, *Chem Phys Lett*, **1971**, *10*, 257 – 261.
- [77] Allen, M. P., Introduction to molecular dynamics simulation, In Attig, N.; Binder, K.; Grubmüller, H.; Kremer, K., Eds., *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes*, Vol. 23 of *NIC*, John von Neumann Institute for Computing, Jülich, Germany, **2004**, 1–28.
- [78] Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H., A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, *J Comput Chem*, **2001**, *22*, 501–508.
- [79] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J Comput Phys*, **1977**, *23*, 327–341.

- [80] Kollman, P., Free energy calculations: applications to chemical and biochemical phenomena, *Chem Rev*, **1993**, *93*, 2395–2417.
- [81] Kästner, J.; Thiel, W., Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “umbrella integration”, *J Chem Phys*, **2005**, *123*, 144104.
- [82] Michalak, A.; Ziegler, T., First-principle molecular dynamic simulations along the intrinsic reaction paths, *J Phys Chem A*, **2001**, *105*, 4333–4343.
- [83] Beveridge, D. L.; DiCapua, F. M., Free energy via molecular simulation: applications to chemical and biomolecular systems, *Annu Rev Biophys Biophys Chem*, **1989**, *18*, 431–492.
- [84] Swegat, W.; Schlitter, J.; Krüger, P.; Wollmer, A., MD simulation of protein-ligand interaction: formation and dissociation of an insulin-phenol complex, *Biophys J*, **2003**, *84*, 1493–1506.
- [85] Ruiz-Montero, M. J.; Frenkel, D.; Brey, J. J., Efficient schemes to compute diffusive barrier crossing rates, *Mol Phys*, **1997**, *90*, 925–942.
- [86] Jarzynski, C., Nonequilibrium equality for free energy differences, *Phys Rev Lett*, **1997**, *78*, 2690–2693.
- [87] Hummer, G., Fast-growth thermodynamic integration: error and efficiency analysis, *J Chem Phys*, **2001**, *114*, 7330–7337.
- [88] Arteca, G. A., Molecular shape descriptors, In Lipkowitz, K. B.; Boyd, D. B., Eds., *Reviews in Computational Chemistry*, Vol. 9, VCH Publishers, New York, **1996**, 191–253.
- [89] Arteca, G. A., Path-integral calculation of the mean number of overcrossings in an entangled polymer network, *J Chem Inf Comput Sci*, **1999**, *39*, 550–557.
- [90] Covey, T. R.; Bonner, R. F.; Shushan, B. I.; Henion, J.; Boyd, R. K., The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry, *Rapid Commun Mass Spectrom*, **1988**, *2*, 249–256.
- [91] Miteva, M.; Demirev, P. A.; Karshikoff, A. D., Multiply-protonated protein ions in the gas phase: calculation of the electrostatic interactions between charged sites, *J Phys Chem B*, **1997**, *101*, 9645–9650.

- [92] Schnier, P. D.; Gross, D. S.; Williams, E. R., On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization, *J Am Soc Mass Spectrom*, **1995**, *6*, 1086 – 1097.
- [93] Gaussian 03, revision c.02. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; Gaussian, Inc., Wallingford CT, **2004**.
- [94] Harrison, A. G., The gas-phase basicities and proton affinities of amino acids and peptides, *Mass Spectrum Rev*, **1997**, *16*, 201–217.
- [95] Wu, Z.; Fenselau, C., Proton affinity of arginine measured by the kinetic approach, *Rapid Commun Mass Spectrom*, **1992**, *6*, 403–405.
- [96] Gogonea, V.; Shy II, J. M.; Biswas, P. K., The electronic structure, ionization potential, and electron affinity of the enzyme cofactor (6R)-5,6,7,8-tetrahydrobiopterin in the gas phase, solution, and protein environments, *J Phys Chem B*, **2006**, *110*, 22861–22871.
- [97] Autenrieth, F.; Tajkhorshid, E.; Baudry, J.; Luthey-Schulten, Z., Classical force field parameters for the heme prosthetic group of cytochrome c, *J Comput Chem*, **2004**, *25*, 1613–1622.
- [98] Liu, D. C.; Nocedal, J., On the limited memory BFGS method for large scale optimization, *Math Program*, **1989**, *45*, 503–528.
- [99] Miyamoto, S.; Kollman, P. A., SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *J Comput Chem*, **1992**, *13*, 952–962.

- [100] Kaltashov, I. A.; Mohimen, A., Estimates of protein surface areas in solution by electrospray ionization mass spectrometry, *Anal Chem*, **2005**, *77*, 5370–5379.
- [101] Sinelnikov, I.; Kitova, E. N.; Klassen, J. S.; Armstrong, G. D., Effects of single amino acid substitution on the dissociation of multiply charged multiprotein complexes in the gas phase, *J Am Soc Mass Spectrom*, **2007**, *18*, 688–692.
- [102] Kohtani, M.; Jones, T. C.; Sudha, R.; Jarrold, M. F., Proton transfer-induced conformational changes and melting in designed peptides in the gas phase, *J Am Chem Soc*, **2006**, *128*, 7193–7197.
- [103] Dongré, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H., Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile proton model, *J Am Chem Soc*, **1996**, *118*, 8365–8374.
- [104] Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brei, L. A., Mobile and localized protons: a framework for understanding peptide dissociation, *J Mass Spectrom*, **2000**, *35*, 1399–1406.
- [105] Vaisar, T.; Urban, J., Gas-phase fragmentation of protonated mono-n-methylated peptides. Analogy with solution-phase acid-catalyzed hydrolysis, *J Mass Spectrom*, **1998**, *33*, 505–524.
- [106] Marinica, D. C.; Grègoire, G.; Desfrancois, C.; Schermann, J. P.; Borgis, D.; Gaigeot, M. P., Ab initio molecular dynamics of protonated dialanine and comparison to infrared multiphoton dissociation experiments, *J Phys Chem A*, **2006**, *110*, 8802–8810.
- [107] Kulhánek, P.; Schlag, E. W.; Koča, J., A novel mechanism of proton transfer in protonated peptides, *J Am Chem Soc*, **2003**, *125*, 13678–13679.
- [108] Paizs, B.; Csonka, I. P.; Lendvay, G.; Suhai, S., Proton mobility in protonated glycylglycine and N-formylglycylglycinamide: a combined quantum chemical and RKKM study, *Rapid Commun Mass Spectrom*, **2001**, *15*, 637–650.
- [109] Neuman, K. C.; Nagy, A., Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy, *Nat Meth*, **2008**, *5*, 491–505.

Appendix A

Charge Sites

1ALA	2GLY	3LEU	4SER	5PRO	6GLU	7GLU	8GLN
9ILE	10GLU	11THR	12ARG	13GLN	14ALA	15GLY	16TYR
17GLU	18PHE	19MET	20GLY	21TRP	22ASN	23MET	24GLY
25LYS	26ILE	27LYS	28ALA	29ASN	30LEU	31GLU	32GLY
33GLU	34TYR	35ASN	36ALA	37ALA	38GLN	39VAL	40GLU
41ALA	42ALA	43ALA	44ASN	45VAL	46ILE	47ALA	48ALA
49ILE	50ALA	51ASN	52SER	53GLY	54MET	55GLY	56ALA
57LEU	58TYR	59GLY	60PRO	61GLY	62THR	63ASP	64LYS
65ASN	66VAL	67GLY	68ASP	69VAL	70LYS	71THR	72ARG
73VAL	74LYS	75PRO	76GLU	77PHE	78PHE	79GLN	80ASN
81MET	82GLU	83ASP	84VAL	85GLY	86LYS	87ILE	88ALA
89ARG	90GLU	91PHE	92VAL	93GLY	94ALA	95ALA	96ASN
97THR	L98EU	99ALA	100GLU	101VAL	102ALA	103ALA	104THR
105GLY	106GLU	107ALA	108GLU	109ALA	110VAL	111LYS	112THR
113ALA	114PHE	115GLY	116ASP	117VAL	118GLY	119ALA	120ALA
121CYS	122LYS	123SER	124CYS	125HIS	126GLU	127LYS	128TYR
129ARG	130ALA	131LYS					

Figure A.1: Amino acid sequence of the cytochrome c' monomer. Charge sites are represented by bold letters.

Appendix B

OPLS Atom Types for the Heme Prosthetic Group of Cytochrome c'

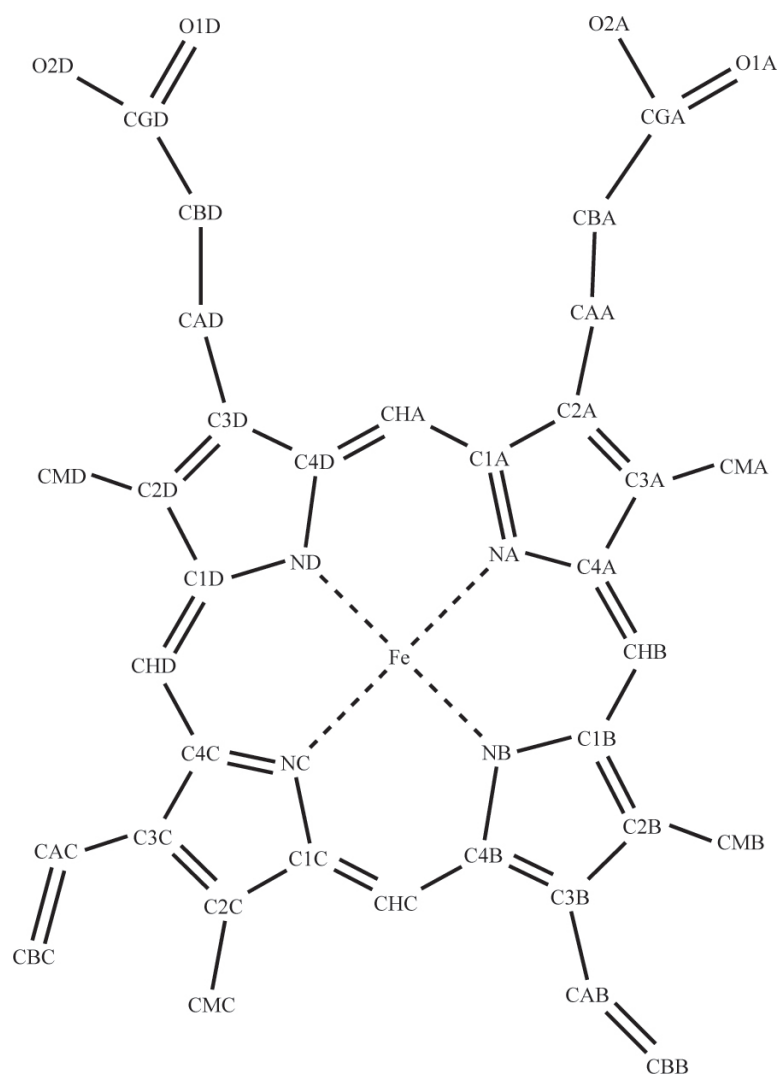


Figure B.1: OPLS atom types for the heme prosthetic group. Hydrogen atoms are also included in parameter file and not shown here for clarity

Appendix C

List of Publications

This thesis is based on the work presented in the following publications:

- 1 Wanasundara, S. N.; Thachuk, M., Theoretical investigations of the dissociation of charged protein complexes in the gas phase, *J. Am. Soc. Mass Spectrom.*, 2007, 18, 2242-2253.
- 2 Wanasundara, S. N.; Thachuk, M., Free energy barrier estimation for the dissociation of charged protein complexes in the gas phase, *J. Phys. Chem. A*, 2009, 113, 3814-3821.
- 3 Wanasundara, S. N.; Thachuk, M., Towards an improved understanding of the dissociation mechanism of gas phase protein complexes, In preparation

Appendix D

Corrections for the Pull Code in Gromacs 3.2

We discovered that the pull code selects the wrong root when it solves the quadratic equation for the COM constraint. This error produces wrong constraint forces. We corrected the error in line 264 of the pull.c file in order to get the correct constraint force.

Line 259 to 269 of initial pull.c file

```
259         if (b < 0)
260             q = 0.5*(b - sqrt(b*b - 4*a*c));
261         else
262             q = 0.5*(b + sqrt(b*b - 4*a*c));
263         x1 = q/a; x2 = c/q;
264         lambda = x1 > 0 ? x1 : x2;
265
266         if (pull->bVerbose)
267             fprintf(stderr, "\nax^2+bx+c=0: a=%e b=%e c=%e\n"
268                     "x1=%e x2=%e sum:%e,%e, lambda:%e\n", a, b, c, x1, x2,
269                     a*x1*x1+b*x1+c, a*x2*x2+b*x2+c, lambda);
```

Corrected code

```
259         if (b < 0)
260             q = 0.5*(b - sqrt(b*b - 4*a*c));
261         else
262             q = 0.5*(b + sqrt(b*b - 4*a*c));
263         x1 = q/a; x2 = c/q;
264         lambda = x1 < 0 ? x1 : x2;
265
266         if (pull->bVerbose)
267             fprintf(stderr, "\nax^2+bx+c=0: a=%e b=%e c=%e\n"
268                     "x1=%e x2=%e sum:%e,%e, lambda:%e\n", a, b, c, x1, x2,
269                     a*x1*x1+b*x1+c, a*x2*x2+b*x2+c, lambda);
```

Note that this error has been fixed in Gromacs 3.2.1 which was released after this thesis work was started.